МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
КИРОВОГРАДСКАЯ ЛЕТНАЯ АКАДЕМИЯ
НАЦИОНАЛЬНОГО АВИАЦИОННОГО УНИВЕРСИТЕТА

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ЭРГОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ

МОНОГРАФИЯ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
КИРОВОГРАДСКАЯ ЛЕТНАЯ АКАДЕМИЯ
НАЦИОНАЛЬНОГО АВИАЦИОННОГО УНИВЕРСИТЕТА

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ЭРГОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ

МОНОГРАФИЯ

Кропивницкий
2016

УДК 331.101.1 : [51-7: 001.891]
ББК Ж30.17-01
М34

Рекомендована к печати ученым сонетом Кировоградской летной академии Национального авиационного университета, протокол от 08.09.2016 № 4

Рецензенты:

В. А. Краснобаев, доктор техн. наук, профессор, Харьков
В. Н. Рудницкий, доктор техн. наук, профессор, Черкассы
А. А. Смирнов, доктор техн. наук, профессор, Кропивницкий

М34 Математические основы эргономических исследований : монография / П. Г. Бердник, Г. А. Кучук, Н. Г. Кучук, Д. Н. Обидин, М.А. Павленко, А.В. Петров, В.Н. Руденко, О.И. Тимочко. – Кропивницкий : КЛА НАУ, 2016. – 248 с.

ISBN 978-966-932-005-6

В монографии обобщены в едином источнике математические методы, которые используются при эргономическом проектировании новой техники с иллюстрацией некоторых аспектов практической применимости затрагиваемых разделов математики. Монографию можно рассматривать как справочник, который поможет спланировать проведение эксперимента, отыскать приемы и способы обоснования решений, разработать модель и исследовать ее. Главной задачей является то, что такую творческую область как эргономика, где большинство решений построены на интуиции, опыте и чувстве прекрасного, можно подкрепить и усилить математикой. Использование математики может позволить формализовать процесс разработки новых систем и разработки новых подходов к проектированию и исследованию.

Для научных сотрудников и специалистов в области эргономики, студентов и аспирантов, изучающих эргономику.

УДК 331.101.1 : [51-7: 001.891]
ББК Ж30.17-01

ISBN 978-966-932-005-6

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В природе существует много такого, что не может быть ни достаточно глубоко понято, ни достаточно убедительно доказано, ни достаточно умело и надёжно использовано на практике без помощи и вмешательства математики.

Фрэнсис Бэкон

Эргономика представляет собой особенную область науки. В этой науке, как ни в какой другой, можно увидеть сплав науки и искусства. Магия науки состоит в озарении ученого, в трудном творческом поиске, основанном на опыте. Путь в эргономике – это путь проб и ошибок, в конце которого лежит огромный клад под названием – опыт. Опыт и накопленные знания позволяют уменьшить количество ошибок, но очень часто этого не достаточно для решения новых задач. Накопленные знания могут сказать о том, чего нельзя делать и ничего о том, как это сделать. Вот в таких случаях и приходит черед озарению или интуиции. Однако без кропотливого и монотонного труда по проверке результатов озарений не мыслим труд ученого-эргономиста. Именно здесь и нужен верный инструмент для проверки полученных результатов. Такими инструментами для эргономиста являются опыт и математика.

Роль математики в любой из наук трудно переоценить. Математика дает инструментарий для исследований во многих областях знаний. Математика - это универсальный язык для понимания и оценки результатов работы других исследователей. Исследования, проводимые в такой области знаний как эргономика, предоставляют огромное количество данных, которое возможно получить, основываясь лишь на эксперименте. Ибо только эксперимент и практика являются лучшими экзаменаторами всей теории, которая разработана в данной отрасли знаний. Вместе с тем, эксперимент - это дорогое удовольствие которое не гарантирует ответа на вопрос: «А можно ли делать именно так?».

В исследованиях несомненно можно опираться на опыт других. Можно интерпретировать знания, полученные другими исследователями, к своим условиям и задачам. Можно использовать свой опыт. Но все эти приемы становятся нереализуемыми для решения новых или впервые появившихся задач. Так, одним из возможных путей решения задачи разработки новых образцов техники в таких условиях становится аналитическое моделирование.

Само понятие моделирования предполагает наличие специфических знаний и навыков для своей реализации. В первую очередь требуются знания о способах и методах построения моделей, исследованиях моделей и обработки знаний, полученных на их основе.

Конечно же, целью данной монографии не является обучение математике тех, кто занимается проблемами, связанными с эргономическим обеспечением процессов создания новых орудий труда. Основная задача данной монографии видится в обобщении в едином источнике математических методов, которые используются при эргономическом проектировании новой техники с иллюстрацией некоторых аспектов практической применимости затрагиваемых разделов математики.

Кроме того, данную монографию можно рассматривать как справочник, который поможет спланировать проведение эксперимента, отыскать приемы и способы обоснования решений, разработать модель и исследовать ее. Главной задачей является то, что такую творческую область как эргономика, где большинство решений построены на интуиции, опыте и чувстве прекрасного, можно подкрепить и усилить математикой. Использование математики может позволить формализовать процесс разработки новых систем и разработки новых подходов к проектированию и исследованию.

Область человеческой деятельности, с которой имеет дело эргономика, неразрывно связана с таким понятием, как система управления. Современное стремительное развитие новых областей знаний влияет и на развитие систем управления, на другие отрасли знаний, с которыми человек сталкивается в процессе своей деятельности. Все это заставляет изменяться и эргономику. Вовлечение новых знаний повышает требование к ученому-эргономисту по овладению новыми знаниями в области математики. Такими областями являются математические методы искусственного интеллекта, исследования операций, системного анализа и другие. Основы знаний в этих областях уже являются необходимыми, а порой и определяющими для получения новых результатов. Повышение уровня интеллектуализации программного обеспечения современных устройств и систем требует абсолютно новых подходов к проектированию среды взаимодействия с ними. Возможность генерирования новых понятий и образов в ходе функционирования систем требует разработки интеллектуальных средств формирования информационной среды работы человека. Становится очевидным тот факт, что мир изменяется все быстрее и эргономике нужно меняться также быстро, чтобы идти в ногу со временем. Перспективное же развитие не возможно без теоретических разработок, построенных с использованием математических моделей и качественной математической проработки возможных вариантов динамики развития процесса.

Данная монография представляет собой первую попытку собрать весь необходимый материал по математике для ученых-эргономистов в одном источнике.

Просьба обо всех замечаниях, пожеланиях и дополнениях сообщать авторам на e-mail: bpgpma@list.ru. Все они будут рассмотрены и необходимые изменения будут внесены в дальнейшие издания.

С уважением, авторы.

ГЛАВА 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Большая часть процессов, с которыми приходится сталкиваться в ходе решения задач эргономики, имеет стохастическую природу. Поэтому первым и основным разделом математики, аппарат которой используется для работы с эргономическими задачами, является теория вероятностей.

1.1. Основные понятия теории вероятностей

1.1.1. Случайные события. Вероятность события

Случайным событием называется любой факт, который в результате проведения опыта может произойти или не произойти.

Достоверным называется событие Ω, которое происходит в каждом опыте.

Невозможным называется событие $\varnothing$, которое в результате опыта произойти не может.

Несовместными называются события, которые в одном опыте не могут произойти одновременно.

Суммой (объединением) двух событий A и B (обозначается $A+B$, $A \cup B$) называется такое событие, которое заключается в том, что происходит хотя бы одно из событий, т.е. A или B, или оба одновременно.

Произведением (пересечением) двух событий A и B (обозначается $A \cdot B$, $A \cap B$) называется такое событие, которое заключается в том, что оба события А и В происходят вместе.

Противоположным событию A называется такое событие $\overline{A}$, которое заключается в том, что событие A не происходит.

События A_k ($k = 1, 2, ..., n$) образуют полную группу событий, если они попарно несовместны и в сумме образуют достоверное событие.

При преобразовании выражений можно пользоваться следующими тождествами:

$$A + \overline{A} = \Omega; \qquad A \cdot \overline{A} = \varnothing; \qquad A \cdot \Omega = A;$$

$$A + \Omega = A; \qquad A \cdot \varnothing = \varnothing; \qquad A + \varnothing = A;$$

$$\overline{A + B} = \overline{A} \cdot \overline{B}; \qquad \overline{A \cdot B} = \overline{A} + \overline{B}; \qquad A + \overline{A} \cdot B = A + B.$$

Классическое определение вероятности. Вероятность случайного события A определяется по формуле:

$$p(A) = \frac{m}{n}, \tag{1.1}$$

где n – число всех равновозможных исходов данного опыта;

m – число равновозможных исходов, приводящих к появлению события.

Геометрическое определение вероятности. Пусть в некоторую область случайным образом попадает точка T, причем все точки области Ω равноправны в отношении попадания точки T. Тогда вероятность попадания точки T в область A определяется отношением:

$$p(A) = \frac{S(A)}{S(\Omega)}, \tag{1.2}$$

где $S(A)$ и $S(\Omega)$ – геометрические меры (длина, площадь, объем и т.д.) областей A и Ω соответственно.

1.1.2. Основные комбинаторные формулы

Пусть имеется множество $X = \{x_1, x_2, ..., x_n\}$, состоящее из n элементов.

Выборкой (n, r) называется множество, состоящее из r элементов, взятых из множества X.

Упорядоченной называется выборка, для которой важен порядок следования элементов. Если каждый элемент множества X может извлекаться несколько раз, то выборка называется выборкой с повторениями.

Число упорядоченных (n, r)-выборок (размещений) с повторениями $\widehat{A}(n,r)$ и без повторений $A(n,r)$ равно:

$$\widehat{A}(n,r) = \mathrm{n}^r, \tag{1.3}$$

$$A(n,r) = \frac{n!}{(n-r)!}. \tag{1.4}$$

Если $r = n$, то размещения без повторений называются перестановками, т.е. расположением элементов исходного множества в определенном порядке.

Число перестановок из n элементов равно:

$$P_n = n! = 1 \cdot 2 \cdot ... \cdot (n-1) \cdot \mathrm{n}\,. \tag{1.5}$$

Пустое множество можно упорядочить только одним способом: $P_0 = 0! = 1$.

Число неупорядоченных (n, r)-выборок (*сочетаний*) с повторениями $\widehat{C}_n^r$ и без повторений C_n^r равно:

$$C_n^r = \frac{(n+r-1)!}{r!(n-1)!}, \tag{1.6}$$

$$C_n^r = \frac{n!}{r!(n-r)!}. \tag{1.7}$$

Число различных разбиений множества из n элементов на k непересекающихся подмножеств (причем в 1-м подмножестве r_1 элементов, во 2-м r_2 элементов и т.д., а $n = r_1 + r_2 + ... + r_k$) равно:

$$P_n(r_1, r_2, ..., r_k) = \frac{n!}{r_1! r_2! ... r_n!}. \tag{1.8}$$

Пример 1.1. В партии транзисторов n стандартных и m бракованных. При контроле оказалось, что первые k транзисторов стандартны. Найти вероятность того, что следующий транзистор будет стандартным.

Решение. Всего осталось для проверки $n + m - k$ транзисторов, из которых стандартных $n - k$. По формуле непосредственного подсчета вероятностей (1.1)

$$p = \frac{n-k}{n+m-k}.$$

Пример 1.2. Банковский сейф имеет кодовый замок, состоящий из шести дисков с восемью буквами на каждом. Сейф открывается при наборе единственной комбинации букв. Злоумышленник пытается открыть сейф, причем на проверку одной кодовой комбинации у него уходит 10 секунд. Какова вероятность того, что злоумышленник успеет открыть сейф, если в его распоряжении 1 час?

Решение. Пусть $P(A)$ – искомая вероятность; n – общее число исходов, равное числу кодовых комбинаций замка, определяемое по формуле (1.3) и равное 8^6; m – число комбинаций, которые успеет испробовать злоумышленник за 1 час, т.е. $m = 360$. Таким образом, искомая вероятность, рассчитывается по формуле (1.1) как m/n:

$$P(A) = \frac{360}{8^6} = 0{,}00137 .$$

1.1.3. Теоремы сложения и умножения вероятностей

Вероятность суммы несовместных событий $A_1, \dots, A_n$ равна сумме вероятностей этих событий:

$$p(A_1 + A_2 + \ldots + A_n) = \sum_{i=1}^{n} p(A_i). \tag{1.9}$$

Вероятность суммы двух совместных событий равна сумме вероятностей каждого из событий минус вероятность их совместного появления:

$$p(A+B) = p(A) + p(B) - p(A \cdot B). \tag{1.10}$$

Вероятность суммы трех совместных событий вычисляется по следующей формуле:

$$\begin{gathered} p(A+B+C) = p(A) + p(B) + p(C) - p(A \cdot B) - \\ - p(B \cdot C) - p(A \cdot C) + p(A \cdot B \cdot C). \end{gathered} \tag{1.11}$$

Вероятность суммы n событий $A_1, \dots, A_n$ равна:

$$\begin{gathered} p\left(\sum_{i=1}^{n} A_i\right) = \sum_{i_1=1}^{n} p(A_{i_1}) - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} p(A_{i_1} A_{i_2}) + \ldots + \\ + (-1)^{k+1} \sum_{i_1=1}^{n} \ldots \sum_{i_k=1}^{n} p(A_{i_1} A_{i_2} \ldots A_{i_k}) + \ldots + (-1)^{n+1} p(A_1 A_2 \ldots A_n). \end{gathered} \tag{1.12}$$

С учетом того, что $p(A) = 1 - p(\overline{A})$, вероятность суммы n событий (если $n > 3$) удобнее вычислять по формуле:

$$p(A_1 + A_2 + \ldots + A_n) = 1 - p(\overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_n}). \tag{1.13}$$

Вероятность произведения 2-х событий равна вероятности одного из них, умноженной на условную вероятность второго при наличии первого

$$p(AB) = p(A)\,p(B/A) = p(B)\,p(A/B). \qquad (1.14)$$

Для независимых событий:

$$p(AB) = p(A)\,p(B). \qquad (1.15)$$

Вероятность произведения n событий $A_i\ (i = 1, 2, ..., n)$ равна:

$$p\,(A_1, A_2, \ldots, A_n) = p(A_1)\cdot p(A_2/A_1)\cdot p(A_3/A_1\cdot A_2)\cdot\ldots\cdot$$
$$\cdot p(A_n/A_1\cdot A_2\cdot\ldots\cdot A_{n-1}), \qquad (1.16)$$

где $p(A_n/A_1\cdot A_2\cdot\ldots\cdot A_{k-1})$ – вероятность появления события A_k, при условии, что события $A_1, A_2, ..., A_{k-1}$ в данном опыте произошли.

В случае независимых событий данная формула упрощается:

$$p\,(A_1, A_2, \ldots, A_n) = p(A_1)\cdot p(A_2)\cdot\ldots\cdot p(A_n). \qquad (1.17)$$

Пример 1.3. Сообщение передается одновременно по n каналам связи, причем для надежности по каждому каналу оно повторяется k раз. При одной передаче сообщение (независимо от других) искажается с вероятностью p. Каждый канал связи (независимо от других) «забивается» помехами с вероятностью q; «забитый» канал не может передавать сообщения. Найти вероятность того, что адресат получит сообщение без искажений.

Решение. Обозначим события:

$$A = \{\textit{хотя бы один раз сообщение передано без искажений}\};$$

$$B_i = \left\{\begin{matrix}\textit{по } i-\textit{му каналу сообщение хотя бы один раз было} \\ \textit{передано без искажений}\end{matrix}\right\}.$$

Для выполнения события B_i i-й канал, во-первых, не должен быть «забит» помехами и, во-вторых, хотя бы одно сообщение по нему не должно быть искажено.

Вероятность того, что канал не «забит» помехами, равна $1-q$.

Вероятность того, что хотя бы одно сообщение передано без помех, равна $1-p^k$ ($p^к$ – вероятность того, что все сообщения переданы с искажениями). Тогда $p(B) = (1-q)\cdot(1-p^k)$.

Вероятность события A, состоящего в том, что хотя бы в одном канале произойдет событие B_i, равна:

$$p(A) = p\left(\bigcup_{i=1}^{n} B_i\right) = 1 - p\left(\overline{A}\right) = 1 - p\left(\bigcap_{i=1}^{n} \overline{B_i}\right) =$$
$$= 1 - \prod_{i=1}^{n}\left(1 - p\left(B_i\right)\right) = 1 - \left[1 - (1-q)\left(1 - p^k\right)\right]^n .$$

1.1.4. Формула полной вероятности. Формула Байеса

Допустим, что проводится некоторый опыт, об условиях которого можно сделать n исключающих друг друга предположений (*гипотез*): $\{H_1, H_2, ..., H_n\}$, $H_i \cap H_j = \varnothing$ при $i \neq j$.

Предположим, что событие A может появляться совместно с одной из гипотез H_i. Тогда полная вероятность события A равна:

$$p(A) = \sum_{i=1}^{n} p(H_i) \cdot p(A / H_i), \tag{1.18}$$

где $P(H_i)$ является априорной вероятностью i-й гипотезы (рассчитанной до появления события *A*).

Если опыт произведен и произошло некоторое событие A, то определить вероятность гипотезы H_k с учетом того, что произошло событие A (апастериорную вероятность данной гипотезы), можно по формуле Байеса:

$$P(H_k / A) = \frac{P(H_k) \cdot P(A / H_k)}{\sum_{i=1}^{n} P(H_i) \cdot P(A / H_i)} . \tag{1.19}$$

Пример 1.4. В продажу поступили телевизоры трех заводов. Продукция первого завода содержит 10% телевизоров с дефектом, второго – 5% и третьего – 3%. Какова вероятность купить неисправный телевизор, если в магазин поступило 25% телевизоров с первого завода, 55% – со второго и 20% – с третьего?

Решение. С рассматриваемым событием A = {приобретенный телевизор оказался с дефектом} связано три гипотезы:

H_1 = {телевизор выпущен первым заводом},

H_2 = {выпущен вторым заводом},

H_3 = {выпущен третьим заводом}.

Вероятности этих событий определяются из условия задачи:

$$p(H_1)=0,25;\ p(H_2)=0,55;\ p(H_3)=0,2.$$

Условные вероятности события A также определяются из условия задачи:

$$p(A/H_1)=0,1;\ p(A/H_2)=0,05;\ p(A/H_3)=0,03.$$

Отсюда по формуле полной вероятности (1.18) ассчитывается вероятность искомого события:

$$p(A)=0,25\cdot 0,1+0,55\cdot 0,05+0,2\cdot 0,03=0,0585$$

1.1.5. Повторение независимых опытов

Пусть производится n независимых одинаковых опытов. В результате каждого опыта событие A появляется с вероятностью p. Вероятность $P(n,k)$ того, что в последовательности из n опытов событие A произойдет ровно k раз (формула Бернулли), равна:

$$P(n,k)=C_n^k\cdot p^k\cdot q^{n-k}=\frac{n!}{k!(n-k)!}p^k\cdot q^{n-k},\ 0\le k\le n, \tag{1.20}$$

где $q=1-p$ – вероятность того, что событие A не произойдет в одном опыте.

Вычисление вероятностей $P(n,k)$ при больших значениях n по формуле Бернулли проблематично.

Поэтому вычисление соответствующих вероятностей проводится с помощью следующих приближенных формул.

Если количество испытаний велико, т.е. $n\to\infty$, а вероятность события мала, т.е. $p\to 0$, так что при этом $np\to a$, $0<a<\infty$ и $p<<\frac{1}{\sqrt{n}}$, то используется формула Пуассона

$$P(n,k)\approx\frac{a^k}{k!}\cdot e^{-a}, k\in\overline{0,n}. \tag{1.21}$$

Если количество испытаний n велико, вероятности p и q не малы, так что выполняются следующие условия:

$$0 < np - 3\sqrt{np}\ ,\ np + 3\sqrt{npq} < 88\ ,\ np + 3\sqrt{npq} < n,$$

то применяются приближенные формулы Муавра–Лапласа:

$$P(n,k) \approx \frac{\phi(x)}{\sqrt{npq}}, \tag{1.22}$$

– локальная $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$,

где $x = \frac{`k - np}{\sqrt{npq}}$;

– интегральная

$$P(n, k_1 \le k \le k_2) \approx \Phi(x_2) - \Phi(x_1), \tag{1.23}$$

где $x_1 = \frac{(k_1 - np)}{\sqrt{npq}}$, $x_2 = \frac{(k_2 - np)}{\sqrt{npq}}$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x \exp\left(-\frac{x^2}{2}\right) dx$ – функция Лапласа.

Функции $\phi(x)$ и $\Phi(x)$ табулированы. При использовании таблиц следует помнить, что $\phi(x)$ является четной $(\phi(-x) = \phi(x))$, а функция Лапласа – нечетной $(\Phi(-x) = -\Phi(x))$.

Пусть производится серия из n независимых испытаний, в результате каждого из которых может появиться одно из событий $A_1, A_2, ..., A_r$ с вероятностями $p_1, p_2, ..., p_r$ соответственно.

Вероятность того, что в серии из n испытаний событие A_1 наступит ровно k_1 раз, событие $A_2 - k_2$ раз, ..., событие $A_r - k_r$ раз $(k_1 + ... + k_r = n)$, равна

$$P(n, k_1, ..., k_r) = \frac{n!}{k_1!...k_r!} \cdot p_1^{k_1} \cdot p_2^{k_2} ... p_к^{k_к}. \tag{1.24}$$

Пример 1.5. По каналу связи передается $n = 6$ сообщений, каждое из которых независимо от других с вероятностью $p = 0,2$ оказывается искаженным. Найти вероятности следующих событий:

A = {ровно два сообщения из шести искажены},

B = {не менее двух сообщений из шести искажены},
C = {все сообщения будут переданы без искажений},
D = {все сообщения будут искажены}.

Решение. По формуле Бернулли (1.20) имеем

$$P(A) = C_6^2 \cdot p^2 \cdot (1-p)^4 = \frac{6!}{4!2!} 0,2^2 \cdot 0,8^4 = 0,197,$$

$$P(B) = P(6,2) + P(6,3) + P(6,4) \pm P(6,5) + P(6,6) = 1 - P(6,0) - P(6,1) =$$

$$= 1 = C_6^0 p^0 (1-p)^6 - C_6^1 p^1 (1-p)^5 = 1 - 0,8^6 - 6 \cdot 0,2^1 \cdot 0,8^5 = 0,345,$$

$$P(C) = (1-p)^6 = 0,262,$$

$$P(D) \equiv p^6 = 0,2^6 = 0,000064.$$

1.1.6. Случайная величина. Закон распределения

Под случайной величиной (*СВ*) понимается величина, которая в результате опыта со случайным исходом принимает то или иное числовое значение, причем заранее, до опыта, неизвестно, какое именно.

Если множество значений *СВ* дискретно, то *СВ* называется дискретной (*ДСВ*).

Закон распределения случайной величины – это любая функция, таблица, правило и т.п., устанавливающая соответствие между значениями случайной величины и вероятностями их наступления.

Функцией распределения случайной величины X называется вероятность того, что она примет значение меньшее, чем аргумент функции x:

$$F(x) = p\{X < x\}. \tag{1.25}$$

Свойства функции распределения:

1. $F(-\infty) = 0$.
2. $F(+\infty) = 1$.
3. При $a < b$

$$p\{x \in [a,\, b]\} = F(b) - F(a). \tag{1.26}$$

Рядом распределения дискретной *СВ* X называется таблица, в верхней строке которой перечислены все возможные, упорядоченные по возрастанию значения *СВ* $x_1, x_2, \ldots, x_n$, а в нижней – вероятности их появления $p_1, p_2, \ldots, p_n$, где $p_i = p\{X = x_i\}$:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Так как события $\{X = x_i\},...\{X = x_n\}$ несовместны и образуют полную группу, то справедливо контрольное соотношение

$$p_1 + p_2 + ... + p_n = 1 . \tag{1.27}$$

Функция распределения любой дискретной CB является разрывной ступенчатой функцией, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений:

$$F(x) = \sum_{x_1 \langle x} p(X = x_1). \tag{1.28}$$

Если у случайной величины функция распределения непрерывна, то она называется непрерывной (*НСВ*).

Плотностью распределения (плотностью вероятности) $f(x)$ непрерывной случайной величины X называется производная ее функции распределения

$$f(x) = \frac{dF(x)}{dx} = F^{'}(x). \tag{1.29}$$

Основные свойства плотности распределения:

1. Плотность распределения неотрицательна: $f(x) \geq 0$.

2. Условие нормировки

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{1.30}$$

3. Вероятность попадания случайной величины X на произвольный участок $[a, b[$ равна

$$p\{a \leq X < b\} = \int_a^b f(x)dx. \tag{1.31}$$

4. Функция распределения $F(x)$ случайной величины X выражается через ее плотность:

$$F(x) = p\{X < x\} = p\{-\infty < X < x\} = \int_{-\infty}^{x} f(x)dx. \qquad (1.32)$$

Пример 1.6. По одной и той же стартовой позиции противника производится пуск из пяти ракет, причем вероятность попадания в цель при каждом пуске одной ракеты равна 0,6. Число попаданий в цель – случайная величина X. Определить ряд распределения и функцию распределения величины X.

Решение. Случайная величина X может принимать следующие значения: 0, 1, 2, 3, 4, 5. Найдем вероятность принятия величиной X этих значений, используя формулу Бернулли:

$P\{X=0\} = (1-p)^5 = 0,4^5 = 0,01024,$

$P\{X=1\} = C_5^1 p(1-p)^4 = 5 \cdot 0,6 \cdot 0,4^4 = 0,0768,$

$P\{X=2\} = C_5^2 p^2(1-p)^3 = 10 \cdot 0,6^2 \cdot 0,4^3 = 0,2304,$

$P\{X=3\} = C_5^3 p^3(1-p)^2 = 10 \cdot 0,6^3 \cdot 0,4^2 = 0,3456,$

$P\{X=4\} = C_5^4 p^4(1-p)^2 = 5 \cdot 0,6^4 \cdot 0,4^2 = 0,2592,$

$P\{X=5\} = p^5 = 0,6^5 = 0,7776.$

Ряд распределения имеет вид

x_i	0	1	2	3	4	5
p_i	0,01024	0,0768	0,2304	0,3456	0,2592	0,07776

Функцию распределения определим по формуле (1.28):

$x \le 0$: $F(x) = 0,$

$0 < x \le 1$: $F(x) = p_0 = 0,01024,$

$1 < x \le 2$: $F(x) = p_0 + p_1 = 0,08704,$

$2 < x \le 3$: $F(x) = p_0 + p_1 + p_2 = 0,31744,$

$3 < x \le 4$: $F(x) = p_0 + p_1 + p_2 + p_3 = 0,66304,$

$4 < x \le 5$: $F(x) = p_0 + p_1 + p_2 + p_3 + p_4 = 0,92224,$

$x > 5$: $F(x) = p_0 + p_1 + p_2 + p_3 + p_4 + p_5 = 1.$

Функцию распределения $F(x)$ можно представить такой таблицей

x	≤ 0	$]0;1]$	$]1;2]$	$]2;3]$	$]3;4]$	$]4;5]$	>5
$F(x)$	0	0,01024	0,08704	0,31744	0,66304	0,92224	1

График искомой функции распределения представлен на рис. 1.1

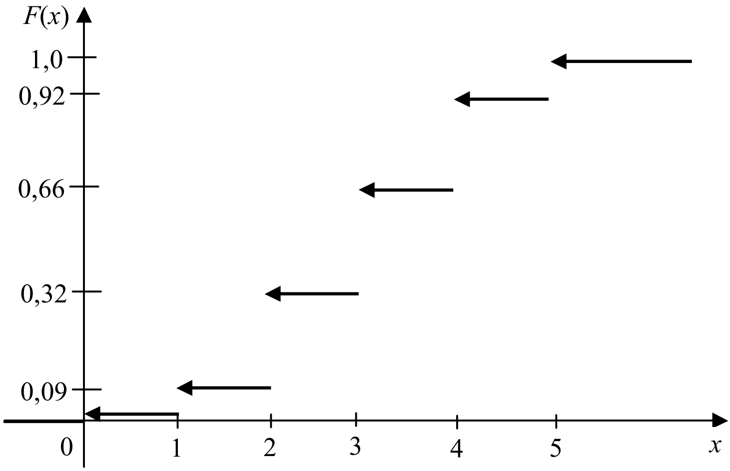


Рис. 1.1. График функции распределения *СВ X*

1.1.7. Числовые характеристики случайной величины

Математическое ожидание характеризует среднее значение *СВ* и определяется по формулам:

$$m_x = M\left[X\right] = \begin{cases} \sum\limits_{i=1}^{N} x_i \cdot p_i & для \;\; ДСВ, \\ \int\limits_{-\infty}^{\infty} x \cdot f(x)dx & для \;\; НСВ. \end{cases} \qquad (1.33)$$

Свойства математического ожидания:

1. $M\left[c\right] = c$. (C – константа или неслучайная величина)

2. $M\left[X + c\right] = M\left[X\right] + c.$

3. $M\left[c \cdot X\right] = c \cdot M\left[X\right].$

Начальным моментом k-го порядка *СВ X* назівается математическое ожидание k-й степени этой случайной величины:

$$a_x = M\left[X^k\right] = \begin{cases} \sum\limits_{i=1}^{N} x_i^k \cdot p_i & \text{для ДСВ}, \\ \int\limits_{-\infty}^{\infty} x^k \cdot f(x)dx & \text{для НСВ}. \end{cases} \tag{1.34}$$

Центрированной случайной величиной X называется CB $\overset{0}{X}$, математическое ожидание которой находится в начале координат (в центре числовой оси абсцисс), т.е. $M\left[\overset{0}{X}\right] = 0$.

Операция центрирования (переход от нецентрированной величины X к центрированной $\overset{0}{X}$) имеет такой вид:

$$\overset{0}{X} = X - m_x.$$

Центральным моментом порядка k CB X назівается математическое ожидание k-й степени центрированной случайной величины $\overset{0}{X}$:

$$\mu_k(x) = M\left[\overset{0}{X}_k\right] = \begin{cases} \sum\limits_{i=1}^{N} (x_i - m_X)^k \cdot p_i & \text{для ДСВ}, \\ \int\limits_{-\infty}^{\infty} (x - m_X)^k \cdot f(x)dx & \text{для НСВ}. \end{cases} \tag{1.35}$$

Второй центральный момент CB называется дисперсией случайной величины, он характеризует степень рассеивания (разброса) значений случайной величины относительно ее математического ожидания и определяется по формулам

$$D_x = D\left[X\right] = \mu_k(x) = a_2(x) - m_X^2 =$$
$$= \begin{cases} \sum\limits_{i=1}^{N} (x_i - m_X)^2 \cdot p_i = \sum\limits_{i=1}^{N} x_i^2 p_i - m_X^2 & \text{для ДСВ}, \\ \int\limits_{-\infty}^{\infty} (x - m_X)^2 \cdot f(x)dx = \int\limits_{-\infty}^{\infty} x^2 f(x)dx - m_X^2 & \text{для НСВ}. \end{cases} \tag{1.36}$$

Для расчетов удобна формула (дисперсия рассчитывается как разность математического ожидания квадрата CB и квадрата числа, равного ее математическому ожиданию):

$$D\left[x\right] = M\left[x^2\right] - \left(m_x\right)^2.$$

Свойства дисперсии:

1. $D[c]=0.$

2. $D[X+c]=D[X].$

3. $D[c\cdot X]=c^2\cdot D[X].$

Средним квадратическим отклонением (СКО) CB X называется число

$$\sigma_x=\sigma[X]=\sqrt{D[X]}. \tag{1.37}$$

СКО измеряется в тех же физических единицах, что и CB, и характеризует ширину диапазона значений CB.

Правило 3σ**.** Большинство значений CB находятся в интервале

$$\left[m_X-3\sigma_X;m_X+3\sigma_X\right]. \tag{1.38}$$

Модой Mo случайной величины называется ее наиболее вероятное значение, т.е. то значение, при котором вероятность p_i (для дискретных CB) или функция $f(x)$ (для непрерывных CB) достигает максимума.

Медианой случайной величины X называется такое ее значение, для которого выполняется условие $p\{X\langle Me\}=p\{X\geq Me\}$.

Медиана, как правило, рассматривается только для непрерывных случайных величин.

Квантилью χ_p случайной величины X называется такое ее значение, для которого выполняется условие $p\{X\langle\chi_p\}=F\{\chi_p\}=p$.

Пример 1.7. Из партии в 25 изделий, среди которых имеется шесть нестандартных, случайным образом выбраны 3 изделия. Найти математическое ожидание и среднее квадратическое отклонение количества нестандартных изделий, содержащихся в выборке.

Решение. По условию задачи CB X принимает следующие значения k: $x_1=0;$ $x_2=1;$ $x_3=2;$ $x_4=3$. Вероятность того, что в этой выборке окажется ровно i $(i=0,\ 1,\ 2,\ 3)$ нестандартных изделий, вычисляется как

$$p_i=P\{X=x_i\}=\frac{C_6^i\cdot C_{19}^{3-i}}{C_{25}^i},$$

следовательно,

$$p_1 = 0,41; \quad p_2 = 0,43; \; p_3 = 0,11; \; p_4 = 0,05.$$

Тогда:

$M\left[X\right] = 0 \cdot 0,41 + 1 \cdot 0,43 + 2 \cdot 0,11 + 3 \cdot 0,05 = 0,8\left(\textit{по формуле } 1.34\right).$

$M\left[X^2\right] a_2(x) = 0^2 \cdot 0,41 + 1^2 \cdot 0,43 + 2^2 \cdot 0,11 + 3^2 \cdot 0,05 = 1,32\left(\textit{по формуле } 1.35\right).$

$D\left[X\right] = 1,32 - (0,8)^2 = 0,68\left(\textit{ по формуле } 1.36\right).$

$\sigma\left[X\right] = 0,82\left(\textit{по формуле } 1.37\right).$

1.1.8. Некоторые часто используемые законы распределения

Дискретная *СВ* X имеет геометрическое распределение, если она принимает значения $0,1,...,\infty$ с вероятностями

$$p(X = i) = p_i = q^i p, \tag{1.39}$$

где p – параметр распределения $(0 \le p \le 1)$, $q = 1 - p$.

Числовые характеристики геометрического распределения:

$$m_X = q / p, D_X = q / p^2.$$

Дискретная *СВ* X имеет биномиальное распределение, если она принимает значения $0,1,..,n$ со следующими вероятностями:

$$p(X = i) = p_i = \frac{n!}{i!(n-i)!} p^i q^{n-i}, \tag{1.40}$$

где n, p – параметры распределения $(0 \le p \le 1), q = 1 - p.$ Числовые характеристики биномиального распределения:

$$m_X = np, \; D_X = nqp.$$

Дискретная *СВ* X имеет распределение Пуассона, если она принимает значения $0,1,...,\infty$ со следующими вероятностями:

$$p(X = i) = p_i = \frac{a^i}{i!} e^{-a}, \tag{1.41}$$

где a – параметр распределения $(a > 0)$.

Числовые характеристики *СВ*, распределенной по закону Пуассона:

$$m_X = a,\ D_X = a.$$

Непрерывная *СВ* X имеет равномерное распределение, если ее плотность вероятности в интервале определения $[a,b]$ постоянна, т.е. если все значения X в этом интервале равновероятны:

$$f(x) = \begin{cases} 0, & x < a, \\ \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & x > b \end{cases} \qquad F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \le x \le b, \\ 0, & x > b \end{cases} \tag{1.42}$$

Числовые характеристики равномерно распределенной *СВ*:

$$m_X = \frac{a+b}{2},\ D_X = \frac{(b-a)^2}{12}.$$

Непрерывная *СВ* T, принимающая только положительные значения, имеет экспоненциальное распределение, если ее плотность вероятности и функция распределения рассчитываются таким образом:

$$f(x) = \begin{cases} \lambda e^{-\lambda t}, & t \ge 0, \\ 0, & t < 0, \end{cases} \quad F(t) = \begin{cases} 1 - e^{-\lambda t}, & t \ge 0, \\ 0, & t < 0, \end{cases} \tag{1.43}$$

где λ – параметр распределения $(\lambda \rangle 0)$.

Числовые характеристики экспоненциально распределенной *СВ*:

$$m_T = 1/\lambda,\ D_T = 1/\lambda^2.$$

Пример 1.8. Время безотказной работы аппаратуры является случайной величиной X, распределенной по экспоненциальному закону. Среднее время безотказной работы 100 ч. Найти вероятность того, что аппаратура проработает больше среднего времени.

Решение. Так как среднее время безотказной работы, т.е. математическое ожидание, равно 100 ч, то параметр λ экспоненциального закона составляет $\lambda = 1/m_X = 1/100 = 0{,}01$. Тогда искомая вероятность

$$p(X \rangle m_X) = p(100 \langle X \langle \infty) = 1 - F(100) = e^{-1} \approx 0{,}368.$$

Непрерывная CB X имеет нормальное распределение, если ее плотность вероятности и функция распределения равны

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\},\ F(x)=0,5+\Phi\left(\frac{x-m}{\sigma}\right), \qquad (1.44)$$

где m,σ – параметры распределения $(\sigma>0)$,

$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ – функция Лапласа

Так как данный интеграл нельзя представить аналитическим выражением, то значения функции Лапласа обычно берутся из таблиц или приближенно рассчитываются после разложения экспоненты в ряд Мак-Лорена. При использовании таблицы значений функции Лапласа следует учитывать, что $\Phi(-x)=-\Phi(x)$, $\Phi(0)=0$, $\Phi(\infty)=0,5$.

Числовые характеристики нормально распределенной CB:

$$m_X=m, D_X=\sigma^2,$$

$$a_k(x)=k!\sum_{i=0}^{I[k/2]}\frac{m^{k-2i}(\sigma/2)^i}{(k-2i)!i!},$$

$$\mu_k(x)=\begin{cases}0, k-\textit{нечетное},\\ \dfrac{k!}{(k/2)!}\left(\dfrac{\sigma^2}{2}\right)^{k/2}, k-\textit{четное}.\end{cases}$$

1.1.9. Функции одного случайного аргумента

Рассмотрим функцию одного случайного аргумента $Y=\phi(X)$. Если X – непрерывная случайная величина, то плотность вероятности $g(y)$ случайной величины Y определяется как

$$g(y)=\sum_{j=1}^{k}f(\psi_j(y))\cdot\left|\psi'_j(y)\right|, \qquad (1.45)$$

где $f(x)$ – плотность вероятности величины X;

$\psi_j(y)$ – функции, обратные для функции $\phi(x)$;

k – число обратных функций для рассматриваемого интервала значений y.

Исходя из (1.45) для нахождения $g(y)$ весь диапазон значений *СВ* Y необходимо разбить на интервалы, в которых число k обратных функций постоянно, и определить вид $g(y)$ для каждого интервала.

Если X – дискретная случайная величина, принимающая значения x_i, то величина Y будет принимать дискретные значения $y_i = \phi(x_i)$ с вероятностями $p(y_i) = p(x_i)$. Числовые характеристики функции $Y = \phi(X)$ одного случайного аргумента X определяются по формулам:

– начальные моменты

$$a_k(y) = M\left[Y^k\right] = M\left[\phi^k(x)\right] = \begin{cases} \sum\limits_{i=1}^{n} \phi^k(x_i)p_i \quad \text{для } \ ДСВ, \\ \int\limits_{-\infty}^{\infty} \phi^k(x)f(x)dx \quad \text{для } \ НСВ; \end{cases} \tag{1.46}$$

– математическое ожидание

$$m_y = M\left[Y\right] = M\left[\phi(x)\right] = a_1(y); \tag{1.47}$$

– центральные моменты

$$\mu_k(y) = M\left[(Y - m_Y)^k\right] = \begin{cases} \sum\limits_{i=1}^{n} (\phi(x_i) - m_y)^k p_i \quad \text{для } \ ДСВ, \\ \int\limits_{-\infty}^{\infty} \phi(x) - m_y)^k f(x)dx \quad \text{для } \ НСВ; \end{cases} \tag{1.48}$$

– дисперсия

$$D_Y = \mu_2(y) = M\left[(Y - m_Y)^2\right] = a_2(y) - m_Y^2. \tag{1.49}$$

Пример 1.8. Определить плотность вероятности величины $Y = X^2$, если X – случайная величина, равномерно распределенная на интервале $\left[-1,\ 2\right]$.

Решение. Так как X равномерно распределена в интервале $\left[-1, 2\right]$, то ее плотность вероятности равна (1.42):

$$f(x) = \begin{cases} 1/3, \ -1 \le x \le 2, \\ 0, \ x < -1, \ x > 2. \end{cases}$$

Рассмотрев функцию $Y = X^2$ для значений $x \in \left[-1,\ 2\right]$ выделим в зависимости от числа k обратных функций такие интервалы для Y:

1) $(-\infty\,,0)$, $k = 0$;

2) $[0\,,1)$, $k = 2$;

3) $[1\,,4)$, $k = 1$;

4) $[4,+\infty)$, $k = 0$.

Так как на интервалах $(-\infty\,,0)$ и $[4,+\infty)$ обратная функция не существует, то для этих интервалов $g(y)=0$.

В интервале $\left[0,\ 1\right]$ – две обратные функции:

$\psi_1(y)=+\sqrt{y}$ *и* $\psi_2(y)=-\sqrt{y}.$

По формуле (1.45) получим

$$g(y)=f_x(\psi_1(y))\cdot\left|\psi_1^{'}(y)\right|+f_x(\psi_2(y))\cdot\left|\psi_2^{'}(y)\right|=$$

$$=f_x(\sqrt{y})\cdot\frac{1}{2\sqrt{y}}+f_x(-\sqrt{y})\cdot\frac{1}{2\sqrt{y}}=\frac{1}{3}\cdot\frac{1}{2\sqrt{y}}+\frac{1}{3}\cdot\frac{1}{2\sqrt{y}}=\frac{1}{3\sqrt{y}}.$$

В интервале $\left]1,4\right]$ – одна обратная функция $\psi_1(y)=+\sqrt{y}$, следовательно,

$$g(y)=f_x(\psi_1(y))\cdot\left|\psi_1^{'}(y)\right|=f_x(\sqrt{y})\cdot\frac{1}{2\sqrt{y}}=\frac{1}{6\sqrt{y}}.$$

Таким образом, плотность вероятности величины Y равна

$$g(y)=\begin{cases}0, & y<0\\ \dfrac{1}{3\sqrt{y}}, & 0\le y\le 1,\\ \dfrac{1}{6\sqrt{y}}, & 1<y\le 4,\\ 0, & y>4.\end{cases}$$

1.1.10. Двумерные случайные величины

Функцией распределения двухмерной случайной величины называется функция двух переменных x и y, определяющая вероятность совместного выполнения двух событий $\{X<x\}$ и $\{Y<y\}$:

$$F(x,y) = p(\{X < x\} \cdot \{Y < y\}). \qquad (1.50)$$

Свойства двумерной функции распределения:

1. $0 \le F(x,y) \le 1.$

2. $F(x,+\infty) = F_X(x);\ \ F(+\infty,y) = F_y(y);\ \ F(+\infty,+\infty) = 1.$

3. $F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0.$

4. $F(x_1,y) \le F(x_2,y),\ \ \textit{если}\ \ x_2 > x_1;$

5. $F(x,y_1) \le F(x,y_2),\ \ \textit{если}\ \ y_2 > y_1.$

Функция распределения может задаваться для любой пары случайных величин.

Для непрерывной двумерной случайной величины (X,Y) существует двумерная плотность распределения:

$$f(x) = \lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}} \frac{p\{x \ge X < x+\Delta x\}\, I\{y \le Y < y+\Delta y\}}{\Delta x \Delta y} = \frac{\partial^2 F(x,y)}{\partial x \partial y}. \qquad (1.51)$$

Свойства двумерной плотности:

1. $f(x,y) \ge 0.$

2. $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy. \qquad (1.52)$

3. $p\{(X,Y) \in D\} = \iint\limits_{(D)} f(x,y)dxdy. \qquad (1.53)$

4. Условие нормировки $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)dxdy = 1. \qquad (1.54)$

5. $f_1(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy;\ \ f_2(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx. \qquad (1.55)$

Для дискретных случайных величин (X,Y) закон распределения задается матрицей распределения, содержащей вероятности p_{ij} появления всех возможных пар значений (x_i, y_j):

$$p_{ij} = p(X = x_i, Y = y_j), \qquad (1.56)$$

удовлетворяющих следующему условию нормировки:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1. \tag{1.57}$$

Одномерные ряды вероятностей, составляющих случайные величины X, Y, определяются по формулам:

$$p_i(X = x_i) = \sum_{j=1}^{m} p_{ij}, i = 1, ..., n; \tag{1.58}$$

$$p_j(Y = y_j) = \sum_{i=1}^{n} p_{ij}, j = 1, ..., m. \tag{1.59}$$

Условным законом распределения называется распределение одной случайной величины, найденное при условии, что другая случайная величина приняла определенное значение.

Условные плотности для непрерывных составляющих X и Y определяются по формулам:

$$f(x/y) = f(x,y)/f_Y(y) \;\; для \;\; f_Y(y) \neq 0; \tag{1.60}$$

$$f(x/y) = f(x,y)/f_X(x) \;\; для \;\; f_X(x) \neq 0. \tag{1.61}$$

Условные ряды распределения для дискретных составляющих X и Y определяются по формулам:

$$p_{i/j} = p(X = x_i / Y = y_j) = p_{ij} / p(Y = y_j), i = 1, ..., N; \tag{1.62}$$

$$p_{j/i} = p(Y = y_i / X = x_i) = p_{ij} / p(X = x_i), j = 1, ..., M \tag{1.63}$$

Случайная величина X независима от случайной величины Y, если ее закон распределения не зависит от того, какое значение приняла случайная величина Y.

Для независимых величин выполняются следующие соотношения:

1. $F(x,y) = p(X<x, Y<y) = p(X<x)p(Y<y) = F_X(x)F_Y(y), \forall x, y;$ (1.64)

2. Для непрерывных величин $- f(x,y) = f_X(x) f_Y(y), \forall x, y.$ (1.65)

3. Для дискретных величин $- p_{ij} = p_i p_j$, для $\forall i, j.$ (1.66)

Пример 1.9. Двумерная случайная величина (X, Y) распределена по закону, приведенному в таблице:

x_i \ y_i	$y_1 = 0$	$y_2 = 1$
$x_1 = -1$	0,1	0,2
$x_2 = 0$	0,2	0,3
$x_3 = 1$	0	0,2

Определить одномерные ряды вероятностей величин X и Y, условный ряд вероятностей величины X при условии, что $Y = 1$. Исследовать зависимость случайных величин X и Y.

Решение. Определим ряды вероятностей X и Y по формулам (1.58) и (1.59), т.е. выполним суммирование по столбцам и по строкам:

y_j	0	1
p_j	0,3	0,7

x_i	-1	0	1
p_i	0,3	0,5	0,2

Условный ряд X при $Y = 1$ получаем по формуле (1.62):

x_i	-1	0	1
$p_i / Y = 1$	2/7	3/7	2/7

Величины X и Y зависимы, так как

$$P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0), 0,2 \neq 0,3 \cdot 0,5 .$$

1.1.11. Числовые характеристики двумерных случайных величин

Рассмотрим основные числовые характеристики двумерной случайной величины (X, Y).

Смешанный начальный момент порядка $k + s$ равен математическому ожиданию произведения X^k и Y^s:

$$a_{k,s}(x, y) = M\left[X^k, Y^s\right] = \begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^k y_j^s p_{i,j} \quad \text{для ДСВ}, \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s f(x, y)dxdy \quad \text{для НСВ}. \end{cases} \tag{1.67}$$

Смешанный центральный момент порядка $k+s$ равен математическому ожиданию произведения центрированных величин $\overset{0}{X}{}^{k}$ и $\overset{0}{Y}{}^{k}$:

$$\mu_{k,s}(x,y) = M\left[(X-m_X)^k (Y-m_Y)^s \right] =$$

$$= \begin{cases} \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} (x-m_x)^k (y_j - m_y)^s \, p_{i,j} \quad \text{для ДСВ}, \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} (x-m_x)^k (y-m_y)^s f(x,y)dxdy \quad \text{для НСВ}. \end{cases} \quad (1.68)$$

где p_{ij} – элементы матрицы вероятностей дискретной величины (X,Y);

$f(x,y)$ – совместная плотность вероятности непрерывной величины (X,Y).

Рассмотрим наиболее часто используемые начальные и центральные моменты:

$$m_X = \alpha_{1?0}(x,y), \, m_Y = \alpha_{0,1}(x,y); \quad (1.69)$$

$$D_X = \mu_{2,0}(x,y) = \alpha_{2,0}(x,y) - m_X^2, \, D_Y = \mu_{0,2}(x,y) = \alpha_{0,2}(x,y) - m_Y^2. \quad (1.70)$$

Корреляционный момент K_{XY} характеризует степень тесноты линейной зависимости величин X и Y и рассеивание относительно точки (m_X, m_Y):

$$K_{XY} = \mu_{1,1}(x,y) = \alpha_{1,1}(x,y) - m_X - m_Y. \quad (1.71)$$

Коэффициент корреляции R_{XY} характеризует степень линейной зависимости величин

$$R_{XY} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{K_{XY}}{\sigma_X \sigma_Y}. \quad (1.72)$$

Для любых случайных величин $|R_{XY}| \le 1$.

Если величины X и Y независимы, то $R_{XY} = 0$.

Пример 1.10. Определить коэффициент корреляции величин X и Y из примера 1.9 предыдущего пункта.

Решение. Определим математические ожидания величин X и Y по формуле (1.69):

$$m_X = \sum_{i=1}^{3} \sum_{j=1}^{2} x_i p_{ij} = \sum_{i=1}^{3} x_i p_i = -1 \cdot 0,3 + 0 \cdot 0,5 + 1 \cdot 0,2 = -0,1;$$

$$m_Y = \sum_{i=1}^{3}\sum_{j=1}^{2} y_j p_{ij} = \sum_{i=1}^{2} y_j p_j = 0\cdot 0,3 + 1\cdot 0,7 = 0,7.$$

Определим $\alpha_{1,1}(x, y)$ по формуле (1.67):

$$\alpha_{1,1}(x, y) = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j p_{ij} = -1\cdot 1\cdot 0,2 + 1\cdot 1\cdot 0,2 = 0.$$

Найдем значение K_{XY} по формуле (1.71):

$$K_{XY} = \alpha_{1,1}(x, y) - m_X m_Y = 0 - (-0,1\cdot 0,7) = 0,07.$$

Определим дисперсии величин X и Y по формуле (1.70):

$$D_X = \alpha_{2,0}(x, y) - m_X^2 = \sum_{i=1}^{3} x_i^2 p_i - m_X^2 = 1\cdot 0,3 + 0\cdot 0,5 + 1\cdot 0,2 - 0,01 = 0,49;$$

$$D_Y = \alpha_{0,2}(x, y) - m_Y^2 = \sum_{j=1}^{2} y_i^2 p_i - m_Y^2 = 0\cdot 0,3 + 1\cdot 0,7 - 0,49 = 0,21.$$

Значение коэффициента корреляции R_{XY} вычислим по формуле (1.72):

$$R_{XY} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{0,07}{\sqrt{0,21\cdot 0,49}} \approx 0,22.$$

1.1.12. Функции случайных величин

Рассмотрим функцию двух непрерывных случайных аргументов $Y = \phi(X_1, X_2)$. Функция распределения $G(y)$ величины Y определяется по формуле

$$G(y) = \iint\limits_{(D)} f(x_1, x_2)dx_1 dx_2, \tag{1.73}$$

где $f(x_1, x_2)$ – совместная плотность вероятности величин X_1 и X_2. Плотность распределения величины Y: $g(y) = G'(y)$.

В формуле (1.73) интегрирование производится по области D, которая определяется из условия $\phi(X_1, X_2)\langle y.$

В случае, когда $Y = X_1, X_2$, функция распределения

$$G(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1} f(x_1, x_2)dx_1 dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_2} f(x_1, x_2)dx_2 dx_1, \tag{1.74}$$

а плотность вероятности

$$g(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1)dx_1 = \int_{-\infty}^{\infty} f(y - x_2, x_2)dx_2. \qquad (1.75)$$

Если величины X_1 и X_2 независимы, то

$$g(y) = \int_{-\infty}^{\infty} f_1(x_1) f(y - x_1)dx_1 = \int_{-\infty}^{\infty} f_1(y - x_2) f_2(x_2)dx_2. \qquad (1.76)$$

Числовые характеристики функции $Y = \phi(X_1, X_2)$ двух случайных непрерывных величин X_1 и X_2, имеющих совместную плотность $f(x_1, x_2)$, определяются по формулам:

– начальные моменты

$$\alpha_k(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi^k(x_1, x_2) f(x_1, x_2)dx_1dx_2; \qquad (1.77)$$

– центральные моменты

$$\mu_k(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\phi(x_1, x_2) - m_y)^k f(x_1, x_2)dx_1dx_2. \qquad (1.78)$$

В случае, когда закон распределения аргументов X_1 и X_2 неизвестен, а известны только их числовые характеристики $m_1, m_2, D_1, D_2, K_{12}$ – математическое ожидание m_Y и дисперсия D_Y величины $Y = X_1 + X_2$ могут быть определены по формулам:

$$m_Y = M\left[X_1 + X_2\right] = m_1 + m_2; \qquad (1.79)$$

$$D_Y = D\left[X_1 + X_2\right] = D_1 + D_2 + 2K_{12}. \qquad (1.80)$$

В случае независимых сомножителей X_1 и X_2 дисперсия $Y = X_1X_2$ может быть определена по формуле

$$D_Y = D\left[X_1 + X_2\right] = D_1D_2 + m_1^2D_2 + m_2^2D_1. \qquad (1.81)$$

Если $Y = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$, α_i неслучайные коэффициенты, то математическое ожидание и дисперсия Y равны:

$$m_Y = M\left[\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right] = \alpha_0 + \sum_{i=1}^{n} \alpha_i m_i; \qquad (1.82)$$

$$D_Y = D\left[\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right] = \sum_{i=1}^{n} \alpha_i^2 D_i + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \alpha_i \alpha_j K_{ij}. \qquad (1.83)$$

Пусть $Y = \prod_{i=1}^{n} X_i$, X_i – независимые случайные величины, значит, математическое ожидание и дисперсия Y равны:

$$m_Y = M\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} m_i; \qquad (1.84)$$

$$D_Y = D\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n}(D_i + m_i^2) - \prod_{i=1}^{n} m_i^2. \qquad (1.85)$$

Пример 1.11. Устройство состоит из двух блоков – основного и резервного. При отказе основного блока автоматически включается резервный блок. Определить вероятность безотказной работы устройства в течение 10 часов, если время безотказной работы блоков случайно и распределено по показательному закону, а среднее время наработки на отказ – 10 часов.

Решение. Определим закон распределения вероятностей времени Y – безотказной работы устройства:

$$Y = X_1 + X_2,$$

где X_1, X_2 – время безотказной работы каждого из блоков.

Величины X_1, X_2 независимы и имеют одинаковую плотность вероятностей:

$$f_1(x) = f_2(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, x\langle 0. \end{cases}$$

Вычислим величину λ. Для показательного закона $\lambda = 1/m_X = 0{,}1$. Определим плотность вероятности Y по формуле (1.76):

$$g(y) = \int_0^y \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda(y-x_1)} dx_1 = \lambda^2 y e^{-\lambda y},\ y>0.$$

Вычислим вероятность того, что $Y>10$:

$$p(Y \geq 10) = \int_{10}^{\infty} g(y)dy = \lambda^2 \int_{10}^{\infty} y e^{-\lambda y} dy \approx 0{,}736.$$

1.2. Математическая статистика

Математическая статистика – раздел математики, областью которой являются математические методы систематизации и использования статистических данных для научных и практических выводов. Во многих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надежность и точность выводов, делаемых на основании ограниченного статистического материала (например, оценить необходимый объем выборки для получения результатов требуемой точности при выборочном обследовании).

1.2.1. Оценка закона распределения

Рассмотрим любую случайную величину *X*. Дадим несколько определений.

Вариант – значение случайной величины, полученное в результате опыта.

Частота варианта или его *вес* – число, показывающее, сколько раз встретился данный вариант в результате нескольких экспериментов.

Частость или *относительная частота* - это отношение веса варианта к числу опытов.

Вариационный ряд – ранжированный в порядке убывания или возрастания ряд вариантов с соответствующими им весами.

Накопленные частости – результат последовательного суммирования частостей в вариационном ряду.

Прерывный вариационный ряд – это вариационный ряд, в котором значения *СВ* отличаются друг от другу не менее, чем на некоторую постоянную величину.

Если множество значений $\{x_1, x_2, ..., x_m\}$ анализируемой *СВ X* конечно, то прерывный вариационный ряд можно представить таблицей:

x_j	x_1	x_2	…	x_m
k_j	k_1	k_2	…	k_m

В данной таблице: $\sum_{j=1}^{m} k_j = n$ – число опытов, k_j –число опытов, в которых получено значение x_j.

Частоту появления j -го значения рассчитывают как $p_j^* = \dfrac{k_j}{n}$.

Непрерывный (или интервальный) вариационный ряд – это вариационный ряд, в котором значения *СВ* могут отличиться одно от другого на сколь угодно малую величину.

В непрерывном вариационном ряду варианты задаются интервалами значений, разности между границами интервала называются *интервальными разностями*.

Пример 1.12. СВ – оценка студента университета на экзамене по математической статистике (100 сдающих, пятибалльная система).

Варианты – x_i	2	3	4	5
Вес (частота) – n_i	10	40	35	15

Построен прерывный вариационный ряд.

Пример 1.13. СВ – процент готовности рабочего места оператора к вводу в эксплуатацию (анализируется 100 рабочих мест).

Варианты – x_i	80 – 100%	60-80%	40 – 60%	20 – 40%	0 – 20%
Вес (частота) – n_i	40	20	20	10	10

Построен непрерывный вариационный ряд (по убыванию числовых значений вариантов), интервальная разность – const, 20%.

Эмпирическая функция распределения вариационного ряда – это функция $F^*(x)$, выражающая для каждого x долю тех элементов, у которых значение случайной величины меньше x, т.е.

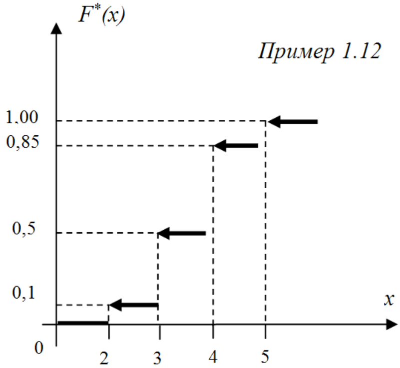
$$F^*(x) = \frac{m(x)}{n}, \tag{1.86}$$

где $m(x)$ – число членов вариационного ряда, имеющих значение *СВ* X, меньшее, чем x.

Значениями эмпирической функции распределения являются накопленные частости, т.е. $F^*(x)$ можно представить в следующем виде:

$$F^*(x) = \begin{cases} 0, & \text{при} \quad x \le x_1; \\ n_1 / n, & \text{при} \quad x_1 < x \le x_2; \\ (n_1 + n_2) / n, & \text{при} \quad x_2 < x \le x_3; \\ \cdots & \\ \sum_{i=1}^{m-1} n_i / n, & \text{при} \quad x_{m-1} < x \le x_m; \\ 1, & \text{при} \quad x > x_m. \end{cases} \tag{1.87}$$

График эмпирической функции распределения прерывного вариационного ряда строится по значениям вариантов, а непрерывного – по серединам интервалов значений. На рис. 1.2 приведены графики $F^*(x)$ для вариационных рядов из примеров 1.12 и 1.13.



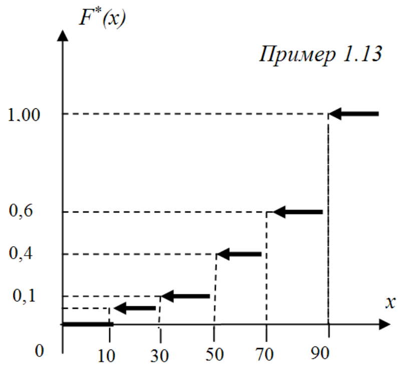


Рис. 1.2. Графики эмпирической функции распределения

Объединив на графиках отрезками точки, являющиеся серединами интервалов, на которых $F^*(x) = const$, получим ломаную линию, которая является довольно хорошим приближением графика функции распределения исследуемой случайной величины (в общем случае – непрерывной).

Заметим, что при большом количестве изучаемых объектов или в случае, когда эксперимент уничтожает объект (например, проверка на надежность) становится проблематично или совсем невозможно составить полный вариационный ряд (провести сплошное наблюдение). Тогда по результатам изучения сравнительно небольшой части совокупности пытаются с достаточной достоверностью получить необходимую информацию о всей совокупности. Такой метод очень часто применяется в практике и называется выборочным. Рассмотрим основные определения, используемые в данном методе.

Генеральная совокупность – это вся совокупность объектов, подлежащих изучению.

Выборка или *выборочная совокупность* – часть генеральной совокупности $\{x_1, x_2, ..., x_n\}$, подлежащая изучению для оценки вероятностных свойств генеральной совокупности.

Объем выборки или генеральной совокупности – это соответствующее число входящих в них объектов.

Основное требование к выборке – выборка должна быть случайна для того, чтобы ее вероятностные свойства были близки к аналогичным свойствам генеральной совокупности. Основными способы, используемыми для этого, являются выбор наугад или использование датчиков (таблиц) случайных чисел.

Выборка, при которой каждый элемент генеральной совокупности может попасть в нее с одинаковой вероятностью, называется *собственно-случайной выборкой.*

Повторная выборка – собственно-случайная выборка, в которой после проверки нужного признака выбранный элемент возвращается.

*Бесповторная выборк*а – собственно-случайная выборка, в которой после проверки нужного признака выбранный элемент обратно не возвращается.

В дальнейшем, если это специально не оговорено, будем рассматривать только собственно-случайные выборки.

Замечание. Распределение признака у выборки – всегда известно, а у генеральной совокупности – всегда неизвестно.

Ошибка репрезентативности – расхождение значений признака в выборочной и генеральной совокупности из-за исследования части совокупности.

Ошибка регистрации – расхождение между истинным и зарегистрированным значениями признака (ошибки измерений и тому подобное).

Чем больше объем выборки, тем меньше ошибка репрезентативности и больше вероятность ошибки регистрации (снижается точность).

Эмпирическая функция распределения $F^*(x)$ вариационного ряда, образованного выборкой, является наилучшей оценкой функции распределения $F(x)$ генеральной совокупности.

1.2.2. Точечные оценки числовых характеристик и параметров

Статистической оценкой $\widehat{Q}$ параметра Q распределения называется приближенное значение параметра, вычисленное по результатам эксперимента (по выборке).

Точечной называется оценка, определяемая одним числом.

Оценка $\widehat{Q}$ называется *состоятельной*, если при увеличении объема выборки n она сходится по вероятности к значению параметра Q:

$$\widehat{Q}(n)\xrightarrow[n\to\infty]{p} Q \Rightarrow \lim_{n\to\infty}\left(P\left(\left|\widehat{Q}(n)-Q\right|<\varepsilon\right)\right)=1 \quad \forall\varepsilon>0 \qquad (1.88)$$

Оценка $\widehat{Q}$ называется *несмещенной*, если ее математическое ожидание точно равно параметру Q для любого объема выборки:

$$\mathrm{M}\left[\widehat{Q}(n)\right]=Q \quad \forall n . \tag{1/89}$$

Несмещенная оценка $\widehat{Q}$ является *эффективной*, если ее дисперсия минимальна по отношению к дисперсии любой другой оценки этого параметра.

Выборочное среднее $\overline{x}$, которое является состоятельной несмещенной оценка математического ожидания исследуемой *СВ*, вычисляется по формуле

$$m_X^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i.} \tag{1.90}$$

Числовые характеристики $\overline{x}$:

$$M\left[\overline{x}\right] = m_X, D\left[\overline{x}\right] = \frac{D_x}{n}. \tag{1.91}$$

Состоятельная несмещенная оценка дисперсии исследуемой *СВ* вычисляется по формуле

$$D_X^* = S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{n}{n-1}\overline{x^2}. \tag{1.92}$$

Числовые характеристики S_0^2:

$$M\left[S_0^2\right] = D_X, D\left[S_0^2\right] = \frac{\mu_4(x)}{n} - \frac{n-3}{n(n-1)}D_X^2. \tag{1.93}$$

Состоятельная несмещенная оценка среднего квадратического отклонения исследуемой *СВ* вычисляется по формуле

$$S_0 = \sqrt{S_0^2}. \tag{1.94}$$

Состоятельная оценка начального момента k-го порядка исследуемой *СВ* вычисляется по формуле

$$\widehat{\alpha}_k(x) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i)^k. \tag{1.95}$$

Состоятельная оценка центрального момента k -го порядка равна

$$\widehat{\mu}_k(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^k. \qquad (1.96)$$

Несмещенная состоятельная и эффективная оценка вероятности случайного события A в схеме независимых опытов Бернулли:

$$p^*(A) = \frac{m}{n}, \qquad (1.97)$$

где m – число опытов, в которых произошло событие A;
n – число проведенных опытов.

Числовые характеристики $p^*(A) = p^*$:

$$M\left[p^*\right] = p(A) = p, D\left[p^*\right] = \frac{p(1-p)}{n}.$$

Для вычисления оценок параметров распределения чаще всего применяются методы моментов и максимального правдоподобия.

Суть метода моментов заключается в следующем. Пусть имеется выборка $\{x_1,...,x_n\}$ независимых значений случайной величины с известным законом распределения $f(x,Q_1,...,Q_m)$ и m неизвестными параметрами $Q_1,...,Q_m$. Последовательность вычислений следующая:

1. Вычислить значения m начальных и/или центральных теоретических моментов

$$\alpha_k(x) = M\left[X^k\right], \quad \mu_л(x) = M\left[(X - m_x)^k\right].$$

2. Определить m соответствующих выборочных начальных $\widehat{\alpha}_k(x)$ и/или центральных $\widehat{\mu}_k(x)$ моментов по формулам (1.95),(1.96).

3. Составить и решить относительно неизвестных параметров $Q_1,...,Q_m$ систему из m уравнений, в которых приравниваются теоретические и выборочные моменты. Каждое уравнение имеет вид $\alpha_k(x) = \widehat{\alpha}_k(x)$ или $\mu_k(x) = \widehat{\mu}_k(x)$. Найденные корни являются оценками $\widehat{Q}_1,\ldots,\widehat{Q}_m$ неизвестных параметров.

Замечание. Часть уравнений может содержать начальные моменты, а оставшаяся часть – центральные.

Согласно методу максимального правдоподобия оценки $\widehat{Q}_1,\ldots,\widehat{Q}_m$ получаются из условия максимума по параметрам $Q_1,...,Q_m$ положительной функции правдоподобия $L(x_1,...,x_n,Q_1,...,Q_m)$.

Если случайная величина X непрерывна, а значения x_i независимы, то

$$L(x_1,...,x_n,Q_1,...,Q_m)=\prod_{i=1}^{n} f(x_i,Q_1,...,Q_m).$$

Если случайная величина X дискретна и принимает независимые значения x_i с вероятностями $p(X=x_i)=p(x_i,Q_1,...,Q_m)$, то функция правдоподобия равна

$$L(x_1,...,x_n,Q_1,...,Q_m)=\prod_{i=1}^{n} p_i(x_i,Q_1,...,Q_m).$$

Система уравнений согласно этому методу может записываться в двух видах:

$$\frac{\partial L(x_1,...,x_n,Q_1,...,Q_m)}{\partial Q_i}=0, \qquad i=1,2,...,m$$

или

$$\frac{\partial I_n(x_1,...,x_n,Q_1,...,Q_m)}{\partial Q_i}=0, \qquad i=1,2,...,m\ .$$

Найденные корни выбранной системы уравнений являются оценками $\widehat{Q}_1,\ldots,\widehat{Q}_m$ неизвестных параметров $Q_1,...,Q_m$.

Пример 1.14. Случайная величина X распределена по равномерному закону, т.е.

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & x<a \vee x>b.\end{cases}$$

Необходимо определить оценки параметров α и b.

Решение. Для данного закона распределения определяем теоретические выражения двух (по числу неизвестных параметров) моментов:

$$a_1(x)=m_X=\int_a^b \frac{x}{(b-a)}dx=\frac{a+b}{2}; \qquad \mu_2(x)=D_X=\frac{(b-a)^2}{12}.$$

По исходной выборке определяем оценки этих же моментов $\overline{x}$ и S_0^2 по формулам (1.91) и (1.92) соответственно. Составляем систему их двух уравнений:

$$\begin{cases} \dfrac{a+b}{2} = \overline{x}, \\ \sqrt{D_X} = \dfrac{b-a}{2\sqrt{3 \cdot S_0.}} \end{cases}$$

Решив ее относительно неизвестных параметров α и b, получим оценки:

$$\hat{a} = \overline{x} - \sqrt{3 \cdot S_0}; \quad \overline{b} = \overline{x} + \sqrt{3 \cdot S_0} .$$

1.2.3. Интервальные оценки числовых характеристик

Оценки $\widehat{Q}$.неизвестного параметра Q определяют одно значение, одну точку на числовой оси, поэтому их называют точечными оценками. Все точечные оценки параметров распределения генеральной совокупности рассчитывают, исходя из выборок, но так как выборки случайны, то и оценки $\widehat{Q}(n)$. являются случайными величинами, отличающимися от постоянного значения истинного параметра Q. Следовательно, можно говорить о точности оценки или об ее предельной ошибке, которую обозначим как $\Delta>0$. Понятно, что чем меньше значение Δ, тем более точна оценка.

Доверительная вероятность γ – это вероятность того, что абсолютная величина отклонения выборочной средней от генеральной средней не превзойдет заданной точности $\Delta > 0$ (предельной ошибкой выборки), т.е.

$$\gamma = P(\left|Q - \widehat{Q}(n)\right| \le \Delta). \tag{1.98}$$

Доверительная вероятность γ задается обычно значением, близким к единице, например, 0,95; 0,98; 0,99 и т. д.

Преобразуем выражение (1.98):

$$\gamma = P\left(-\Delta \le Q - \widehat{Q}(n) \le \Delta\right) = \gamma = P\left(\widehat{Q}(n) - \Delta \le Q \le \widehat{Q}(n) + \Delta\right).$$

Из полученного выражения видно, что интервал длиной 2Δ накрывает значение параметра Q с заданной доверительной вероятностью γ.

Доверительный интервал оценки параметра Q – это интервал длиной 2Δ (двойная предельная ошибка выборки) с центром, равным значению оценки.

Связь между предельной ошибкой выборки Δ, доверительной вероятностью γ и средней ошибкой выборки $\bar{\sigma}$ устанавливает следующее утверждение: если $\Phi(t) = \gamma$, то предельная ошибка выборки равна произведению числа t на среднюю ошибку выборки, т.е.

$$\Delta = t \cdot \bar{\sigma} \,. \tag{1.99}$$

Исходя из (1.99) предельную ошибку выборки можно рассчитать по таким формулам (σ – СКО выборки):

$\Delta = t \cdot \sigma / \sqrt{n}$ (для выборочной средней повторной выборки);

$\Delta = t \cdot \sigma / \sqrt{n} \cdot \sqrt{1 - n / N}$ (для выборочной средней бесповторной выборки).

Рассмотрим нахождение доверительных интервалов для основных параметров нормального распределения: математического ожидания μ и дисперсии σ^2. Предположим, что задана генеральная совокупность с нормальным распределением, $X \in N(\mu, \sigma)$. Рассмотрим несколько случаев.

1. *Доверительный интервал для μ при известном σ.* В данном случае считаем, что $\sigma_0 \approx \sigma$, а σ получено в результате дополнительной выборки или известно из предыдущих экспериментов. Заметим, что и среднее арифметическое выборки $\bar{x}$, и элементы выборки $x_1, x_2, \ldots, x_n$ из-за случайности выборок являются случайными величинами. Все элементы выборки имеют тоже распределение, что и генеральная совокупность, следовательно $x_i \in N(\mu, \sigma)$. Среднее арифметическое выборки также имеет нормальное распределение: $\bar{x} \in N\left(\mu,\ \sigma/\sqrt{n}\right)$. Тогда доверительную вероятность γ можно рассчитать следующим образом:

$$\gamma = P\left(\left|\bar{x} - \mu\right| \le \Delta\right) = P\left(\mu - \Delta \le \bar{x} \le \mu + \Delta\right) = \Phi\left(\frac{\mu - \Delta - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu + \Delta - \mu}{\sigma/\sqrt{n}}\right) =$$

$$= \Phi\left(\frac{\Delta\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\Delta\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 = 2\Phi\left(u_\gamma\right) - 1,$$

где $u_\gamma = \sqrt{n} \cdot \Delta / \sigma$; $\Phi(z)$ – функция Лапласа.

Значит $\Delta = u_\gamma \cdot \sigma / \sqrt{n}$, т.е. формула для вычисления доверительного интервала имеет такой вид:

$$P\left(\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}}\right) = \gamma\ . \qquad (1.100)$$

Для вычисления значение переменной u_γ воспользуемся условием

$$\Phi\left(u_\gamma\right) = 1/2\left(1+\gamma\right),$$

согласно которому значение аргумента u_γ находится из таблиц для функции Лапласа.

2. *Доверительный интервал для* μ *при неизвестном* σ. Так как σ неизвестно, то непосредственно воспользоваться нормальным распределением N (μ, σ) нельзя. Однако известно, что случайная величина

$$t = \frac{\bar{x} - \mu}{\sigma}\sqrt{n}\ ,$$

где σ – несмещенная оценка стандартного отклонения генеральной совокупности, имеет распределение Стьюдента (называемое иногда t-распределением) с числом степеней свободы n – 1.

Для получения границ доверительного интервала оценки потребуем, чтобы выполнялось условие

$$P\left(\left|\frac{\bar{x} - \mu}{\sigma}\sqrt{n}\right| \le t_\gamma\right) = P\left(\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}\right) = \gamma\ , \qquad (1.101)$$

т.е. предельная ошибка выборки равна $\Delta = t_\gamma \dfrac{\sigma}{\sqrt{n}}$.

Величина t_γ определяется по таблицам распределения Стьюдента на основании условия P(|X| > t_γ) = γ с числом степеней свободы n – 1, т.е.

$$P\left(\left|\frac{\bar{x} - \mu}{\sigma}\sqrt{n}\right| > t_\gamma\right) = 1 - \gamma\ .$$

3. *Доверительный интервал для дисперсии* S_0^2. Известно, что случайная величина

$$\chi^2 = \frac{(n-1)S^2}{S_0^2}\ ,$$

где S^2 - выборочная дисперсия, имеет χ^2 - распределение (распределение Пирсона) с числом степеней свободы n – 1. По таблицам χ^2 –распределения можно найти пару чисел u_1, u_2, таких, что:

$$P\left(u_1 \le \chi^2 \le u_2\right) = \gamma \text{ ;}$$

$$P\left(\chi^2 < u_1\right) = P\left(\chi^2 > u_2\right) = \frac{1}{2}(1-\gamma) \text{ .}$$

Тогда доверительный интервал дисперсии рассчитывается из выражения

$$P\left(\frac{(n-1)S^2}{u_2} \le S_0^2 \le \frac{(n-1)S^2}{u_1}\right) = \gamma \text{ .} \qquad (1.102)$$

Доверительный интервал с удовлетворительной вероятностью γ для вероятности события A в схеме независимых опытов Бернулли

$$p^* - z_\gamma \cdot \sqrt{\frac{p^*(1-p^*)}{n}} < p(A) < p^* + z_\gamma \cdot \sqrt{\frac{p^*(1-p^*)}{n}}, \qquad (1.103)$$

где $p^* = p^*(A) = \dfrac{m}{n}$ – частота появления события A в n опытах;

m – число опытов, в которых произошло событие A;

n – число проведенных опытов.

Пример 1.15. Производится серия независимых опытов с целью определения вероятности события A. В 100 опытах событие произошло 40 раз. Частота события $p^*(A) = \dfrac{m}{n}$ принимается за приближенное значение вероятности этого события. Найти вероятность того, что допущенная при этом ошибка меньше 0,1.

Решение. Необходимо найти доверительную вероятность для следующего доверительного интервала:

$$p(\left|p^*(A) - p(A)\right| \le 0{,}1) = p(p^* - 0{,}1 < p(A) < p^* + 0{,}1) = \gamma,$$

т.е. $z_\gamma \cdot \sqrt{\dfrac{p^*(1-p^*)}{n}} = 0{,}1$ (из формулы 1.103).

С учетом того, что $p^* = \dfrac{40}{100} = 0{,}4, z_\gamma = \dfrac{0{,}1 \cdot \sqrt{100}}{\sqrt{0{,}4 \cdot 0{,}6}} = 2{,}041$, искомая доверительная вероятность равна

$$\gamma = 2 \cdot \Phi(2{,}041) \approx 0{,}958.$$

Пример 1.16. Стальная проволока исследуется на разрыв. Сделана выборка из экспериментальной партии в 400 проволок объемом 100 проволок со следующим вариационным рядом:

x_i – сила (кг/мм^2)	[40-42)	[42-44)	[44-46)	[46-48)	[48-50)
n_i - к-во проволок	7	24	38	19	12

Найти доверительную вероятность того, что среднее разрывное усилие отличается от выборочной средней, не более, чем на величину 0,4 кг/мм^2 (естественно, для бесповторной выборки).

Решение. Шаг 1. Рассчитываем выборочную среднюю и выборочную дисперсию по данным вариационного ряда:

$$\overline{x} = 45{,}1\,; \quad S^2 = 4{,}75\,.$$

Шаг 2. Рассчитываем среднюю квадратическую ошибку:

$$\overline{\sigma}_б = \sqrt{\frac{4{,}75}{100}\left(1 - \frac{100}{400}\right)} \approx 0{,}19 \text{ (бесповторная выборка).}$$

Шаг 3. Рассчитываем доверительную вероятность:

$$\beta = \mathrm{P}_Б\left(\left|x_0 - 45{,}1\right| \le 0{,}4\right) \approx \Phi\left(0{,}4/0{,}19\right) \approx \Phi\left(2{,}1\right) \approx 0{,}9821\,.$$

1.2.4. Проверка статистических гипотез о законе распределения

Критерием согласия называется случайная величина от выборки

$$U = \varphi\left(x_1, x_2, \ldots x_n\right),$$

где x_i – значения выборки, которая позволяет принять или отклонить гипотезу о предполагаемом законе распределения.

1.2.4.1. *Последовательность действий при проверке гипотезы о законе распределения при помощи критерия согласия* χ^2 :

1. Построить по выборке вариационный ряд гистограмму.
2. По виду гистограммы выдвинуть гипотезы:

$$\text{основная гипотеза } H_0 : F(x) = F_0(x),$$

$$\text{альтернативная гипотеза } H_1 : F(x) \ne F_0(x),$$

где $F_0(x)$ – плотность и функция гипотетического закона распределения.

3. Используя метод моментов или максимального правдоподобия, определить оценки неизвестных параметров $\widehat{Q}_1, \ldots, \widehat{Q}_m$ гипотетического закона распределения.

4. Вычислить значение критерия по формуле

$$\chi^2 = n\sum_{j=1}^{M}\frac{(p_j - p_j^*)^2}{p_j} = \sum_{j=1}^{M}\frac{(v_j - np_j)^2}{np_j}, \tag{1.104}$$

где p_j – теоретическая вероятность попадания случайной величины в j-й интервал при условии, что гипотеза H_0 верна:

$$p_j = p(A_j \le X\langle B_j) = \int_{A_j}^{B_j} f_0(x)dx = F_0(B_j) - F_0(A_j). \tag{1.105}$$

Замечание. При расчете p_1 и p_M в качестве крайних границ первого и последнего интервалов A_1, B_M следует использовать теоретические границы гипотетического закона распределения. Например, для нормального закона $A_1 = -\infty, B_M = +\infty$. После вычисления всех вероятностей p_i проверить, выполняется ли контрольное соотношение

$$\left|1 - \sum_{j=1}^{M} p_i\right| \le 0,01.$$

5. Из таблицы χ^2 выбирается значение $\chi^2_{a,k}$ (граница критической области), где α – заданный уровень значимости (обычно α = 0,05 или 0,01), а k – число степеней свободы, определяемое по формуле

$$k = M - I = s,$$

где s – число параметров гипотетического закона распределения, значения которых были определены в п. 3.

6. Если $\chi^2\rangle\chi^2_{a,k}$, то гипотеза H_0 отклоняется, в противном случае нет оснований ее отклонить.

1.2.4.2. *Последовательность действий при проверке гипотезы о законе распределения при помощи критерия согласия Колмогорова.*

1. Построить вариационный ряд и график эмпирической функции распределения $F^*(x)$.

2. По виду графика $F^*(x)$ выдвинуть гипотезы:

$$\text{основная гипотеза } H_0 : F(x) = F_0(x),$$

$$\text{альтернативная гипотеза } H_1 : F(x) \ne F_0(x),$$

где $F_0(x)$ – функция гипотетического закона распределения.

3. Используя метод моментов или максимального правдоподобия, определить оценки неизвестных параметров $\widehat{Q}_1, \ldots, \widehat{Q}_m$.

4. Рассчитать 10 – 20 значений функции $F_0(x)$ и построить ее график в одной системе координат с функцией $F^*(x)$.

5. По графику определить максимальное по модулю отклонение между функциями $F^*(x)$ и $F_0(x)$.

$$Z = \max_{i=1}^{n} \left| F^*(x_i) - F_0(x_i) \right|. \tag{1.106}$$

6. Вычислить значение критерия Колмогорова

$$\lambda = \sqrt{n \cdot Z}. \tag{1.107}$$

7. Из таблицы распределения Колмогорова выбрать критическое значение $\lambda_\gamma, \gamma = 1 - a$, α – заданный уровень значимости (α = 0,05 или 0,01).

8. Если $\lambda \rangle \lambda_\gamma$, то основная гипотеза H_0 отклоняется, в противном случае нет оснований ее отклонить.

Пример 1.17. По заданной целочисленной выборке при уровне значимости $\alpha = 0{,}01$ проверить гипотезу о распределении Пуассона соответствующей генеральной совокупности ($p_m = \lambda_m e^{-\lambda} / m!$; $\lambda = np = m_x = D_x$).

Элементы выборки 2; 4; 2; 4; 3; 3; 3; 2; 0; 6; 1; 2; 3; 2; 2; 4; 3; 3; 5; 1; 0; 2; 4; 3; 2; 2; 3; 3; 1; 3; 3; 3; 1; 1; 2; 3; 1; 4; 3; 1; 7; 4; 3; 4; 2; 3; 2; 3; 3; 1; 4; 3; 1; 4; 5; 3; 4; 2; 4; 5; 3; 6; 4; 1; 3; 2; 4; 1; 3; 1; 0; 0; 4; 6; 4; 7; 4; 1; 3.

Решение: Шаг 1. Анализируем выборку и строим вариационный ряд:

0	1	2	3	4	5	6	7	*xi*
4	12	14	24	17	3	3	2	частота
0,05	0,16	0,18	0,3	0,2	0,04	0,04	0,03	%

Шаг 2. Строим полигон (рис. 1.3) и эмпирическую функцию распределения (рис. 1.4).

Рис. 1.3. Полигон ВР примера 1.17

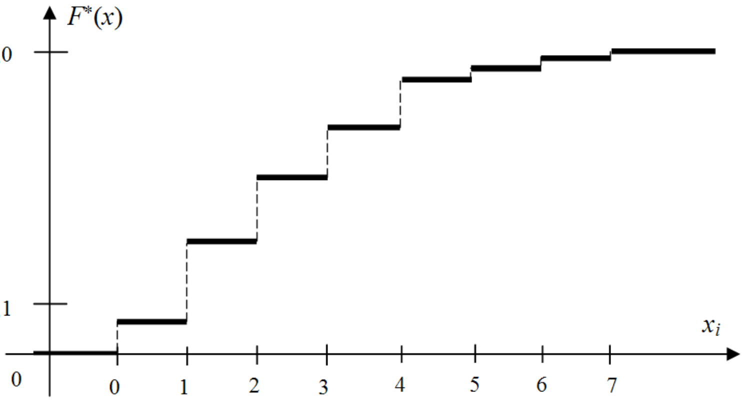


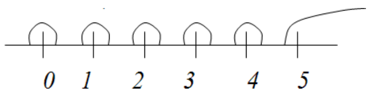
Рис. 1.4. Эмпирическая функция распределения примера 1.17

Шаг 3. Рассчитываем числовые характеристики *ВР*:

$$\overline{x} \approx 2{,}84\,; \quad S^2 \approx 2{,}4; \quad n = 79.$$

Шаг 4. Так как для распределения Пуассона $m_x = D_x = \lambda$, то будем проверять только один параметр – параметр распределения λ. Заметим, что выборочные оценки m_x и D_x относительно близки, но не равны. Для упрощения расчетов предположим, что параметр распределения является целым числом. Наиболее близким к оценкам целым числом является значение $\lambda = 3$.

Шаг 5. Разобьем область значений *СВ X* исходя из построенного вариационного ряда на 6 интервалов ($k = 6$):



Шаг 6. Статистика χ^2 имеет χ^2-распределение с $k - r - 1 = 4$ степенями свободы. Исходя из условия задачи и таблиц χ^2-распределения:

$$\mathrm{P}(\chi^2 > \chi^2_{\alpha,k}) = 0{,}01 \quad \Rightarrow \quad \chi^2_{\alpha,k} \approx 13{,}3.$$

Шаг 7. Рассчитываем значение статистики χ^2 для рассматриваемого вариационного ряда. Для этого можем воспользоваться табл. 1.1, в которой p_i – теоретические вероятности, рассчитываемые по формулам распределения Пуассона $p_k = \left(\lambda^k / k!\right) \cdot e^{-\lambda}$ или найденные по таблицам распределения для целых параметров; m_i – эмпирические вероятности.

Таблица 1.1

Расчет статистики заданного *ВР*

x_i	n_i	p_i	$m_i = n \cdot p_i$	$n_i - m_i$	$(n_i - m_i)^2$	$(n_i - m_i)^2 m_i$
0	4	0,05	3,9	0,1	0,01	0,0026
1	13	0,15	11,8	1,2	1,44	0,1220
2	14	0,225	17,7	-3,7	13,69	0,7734
3	24	0,225	17,7	6,3	39,69	2,2424
4	14	0,17	13,3	2,7	7,29	0,5481
≥ 5	8	0,18	14,6	-6,6	43,56	2,9836
Σ	79	1	79			$\chi^2 \approx 6{,}67$

Шаг 7. Так как полученное значение статистики $\chi^2 \approx 6{,}67$ меньше границы критической области $\chi^2_{\alpha,k} \approx 13{,}3$, то принимается основная гипотеза H_0: генеральная совокупность имеет распределение Пуассона.

1.2.5. Оценка коэффициента корреляции и линейной регрессии

Пусть проводится n независимых опытов, в каждом из которых двумерную *СВ* (X, Y) принимает определенные значения и результаты опытов представляют собой двухмерную выборку вида $\{x_i,\ y_i\}$. Первичная обработка опытных данных включает в себя обработку составляющих X и Y как одномерных величин и вычисление оценок, присущих только двухмерным (многомерным) случайным величинам.

Между случайными величинами X и Y могут существовать две различные зависимости:

между значениями x и средними значениями $\overline{y}$ (регрессия Y на X);

между значениями y и средними значениями $\overline{x}$ (регрессия X на Y).

Уравнения регрессии – это уравнения, выражающие в общем виде данные зависимости, а кривые регрессии – это их графики:

$$\begin{aligned} f(x) &= M\left(Y \middle| X = x\right); \\ g(y) &= M\left(X \middle| Y = y\right). \end{aligned} \tag{1.108}$$

Эмпирическая регрессия – регрессия, полученная в результате наблюдений.

Ковариация или корреляционный момент СВ (X, Y) определяется через центральный момент следующим образом:

$$K_{xy} = \operatorname{cov}(X,Y) = \mu_{1,1} == M\left[\overset{\circ}{X} \cdot \overset{\circ}{Y}\right] =$$
$$M\left[\left(X - m_x\right)\left(Y - m_y\right)\right] = M[XY] - M[X] \cdot M[Y]. \qquad (1.109)$$

Корреляционный момент в математической статистике – это мера линейной зависимости случайных величин.

Коэффициент корреляции СВ (X,Y) определяется так:

$$R_{xy} = \frac{K_{xy}}{\sigma_x \cdot \sigma_y} \qquad (1.110)$$

Коэффициент корреляции в математической статистике – это нормированная мера линейной зависимости случайных величин. Так как корреляционный момент (в отличие от коэффициента корреляции) не инвариантен относительно смены масштаба, что не всегда удобно в приложениях, то более часто при анализе зависимости двух случайных величин используют значение коэффициента корреляции.

Свойства корреляционного момента:

1) симметричность, т.е. $K_{xy} = K_{yx}$;

2) связь с дисперсией: $D_x = M\left[\overset{\circ}{X}\,\overset{\circ}{X}\right] = K_{xx}$; $D_y = M\left[\overset{\circ}{Y}\,\overset{\circ}{Y}\right] = K_{yy}$.

3) для независимых случайных величин корреляционный момент равен нулю (обратное, вообще говоря, неверно);

4) корреляционный момент характеризует как степень зависимости *СВ* X и Y, так и их рассеивание вокруг точки (m_x, m_y);

5) неравенство Коши – Буняковского:

$$K_{xy} \le D\left[X\right] \cdot D\left[Y\right] \qquad (1.111)$$

Свойства коэффициента корреляции

1) абсолютная величина коэффициента корреляции не превосходит единицы, т.е. $\left|R_{xy}\right| \le 1$;

2) равенство модуля коэффициента корреляции единице – это необходимое и достаточное условие того, что *СВ* Y и X связаны линейной функциональной зависимостью;

3) если коэффициент корреляции равен нулю, то между y и x нет линейной корреляционной связи., при этом нелинейная связь может быть;

4) выполнение условия $\left|R_{xy}\right| = 1$ является необходимым и достаточным условием совпадения прямых регрессии.

Перейдем к рассмотрению оценок коэффициента корреляции и линейной регрессии.

Состоятельная несмещенная оценка корреляционного момента равна

$$K_{XY}^{*} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}), \tag{1.112}$$

где x_i, y_i – значения, которые приняли случайные величины X, Y в i-м опыте; $\overline{x}$, $\overline{y}$ – средние значения случайных величин X и Y соответственно.

Состоятельная оценка коэффициента корреляции

$$R_{XY}^{*} = \frac{K_{XY}^{*}}{S_0(x) S_0(y)}. \tag{1.113}$$

Доверительный интервал с надежностью γ для коэффициента корреляции R_{XY}^{*} и случая двухмерного нормального распределения

$$\frac{e^{2a}-1}{e^{2a}+1} < R_{XY} < \frac{e^{2b}-1}{e^{2b}+1}, \tag{1.114}$$

где значения a и b определяются исходя из заданной надежности (доверительной вероятности) по таблицам функции Лапласа.

Рассмотрим алгоритм проверки гипотезы об отсутствии корреляционной зависимости (предполагается, что двухмерная случайная величина (X, Y) распределена по нормальному закону).

1. Формулируются основная и альтернативная гипотезы: $H_0 : R_{XY} = 0$; $H_1 : R_{XY} \neq 0$, где R_{XY} – теоретический коэффициент корреляции.

2. Вычисляется оценка коэффициента корреляции R_{XY}^{*} по формуле (1.113).

3. Определяется значение критерия

$$t = \frac{R_{XY}^{*} \sqrt{n-2}}{\sqrt{1-(R_{XY}^{*})}}, \tag{1.115}$$

который распределен по закону Стьюдента с $(n-2)$ степенями свободы.

4. По заданному уровню значимости α вычисляется доверительная вероятность $\gamma = 1 - \alpha$ и из таблицы распределения Стьюдента выбирается критическое значение $t_{\gamma, n-2}$.

5.Если $|t| > |t_{\gamma,n-2}|$, то гипотеза H_0 отклоняется, а, следовательно, величины X, Y коррелированны, иначе гипотеза H_0 принимается.

Регрессию случайной величины Y на X можно определить используя условное математическое ожидание случайной величины Y при условии, что $X = x$:

$$m_{Y/x} = M\left[Y / X = x\right]. \tag{1.116}$$

Регрессия Y на X устанавливает зависимость среднего значения величины Y от величины X. Если X и Y независимы, то

$$m_{Y/x} = m_Y = const.$$

Представляет интерес линейная регрессия, для которой

$$m_{Y/x} = \alpha_0 + \alpha_1 x. \tag{1.117}$$

Оценки коэффициентов α_0 и α_1 из формулы (1.117) по методу наименьших квадратов вычисляются так:

$$\widehat{\alpha}_1 = K^*_{XY} \big/ S_0^2\left(x\right); \quad \widehat{\alpha}_0 = \overline{y} - \alpha_0 \cdot \overline{x}, \tag{1.118}$$

где $\overline{x}, \overline{y}$ - оценки математического ожидания *СВ* X, Y;

$S_0^2(x)$ - оценка дисперсии *СВ* X;

K^*_{XY} – оценки корреляционного момента *СВ* X и Y.

Для визуальной проверки правильности вычисления этих оценок можно построить диаграмму рассеивания и график прямой $\overline{y}\left(x\right) = \widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot x$ (рис. 1.5).

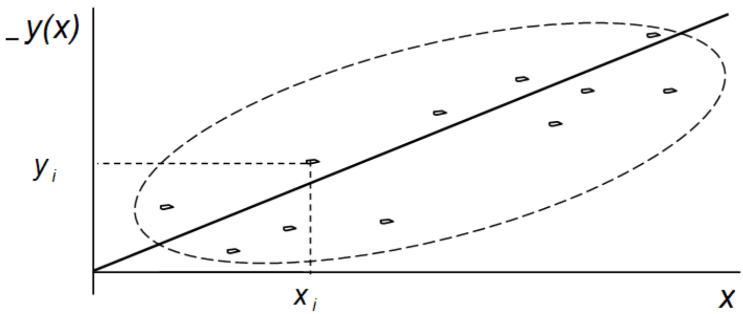


Рис. 1.5. Графический анализ оценки линейности регрессии

Если оценки параметров a_0, a_1 рассчитаны без грубых ошибок, то сумма квадратов отклонений всех точек (x_i, y_i) от прямой $\overline{y}(x) = \widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot x$ должна быть минимально возможной.

1.2.6. Методы математической статистики в эргономике

Методы математической статистики широко применяются в эргономике.

Ниже приведен пример решения задачи проверки согласия эмпирических и теоретических распределений затрат времени на выполнение действий оператором.

При моделировании деятельности операторов одним из основных вопросов является вопрос выбора закона распределения затрат времени на выполнение элементарных операций. При этом необходимо учитывать, что затраты времени имеют минимальное и максимальное значение и при каждой реализации модели имеют случайные значения. Обоснование выбора закона распределения затрат времени является сложной задачей, которую необходимо решать при моделировании деятельности каждого оператора в отдельности.

Основой для решения данной задачи является множество значений времени по решению конкретных задач реальными операторами, полученными в ходе проведения экспериментов либо замеров в процессе реальной работы операторов.

В табл. 1.2 и на рис. 1.6 показаны данные по задаче набора цифровых символов и гистограмма и плотности распределения времени решения оператором данной задачи.

Таблица 1.2

Ввод команд оператором

№№ п\п	Количество набираемых цифр	Значения χ^2			
		Усеч. норм.	Логарифм. норм.	Бета	Гамма
1	1	43	627	12,9	36,5
2	2	11,7	21,3	9,8	460
3	3	6,7	36,6	21,0	1023
4	4	25,7	149	7,7	1023
5	5	20,1	88.5	7,2	52.6

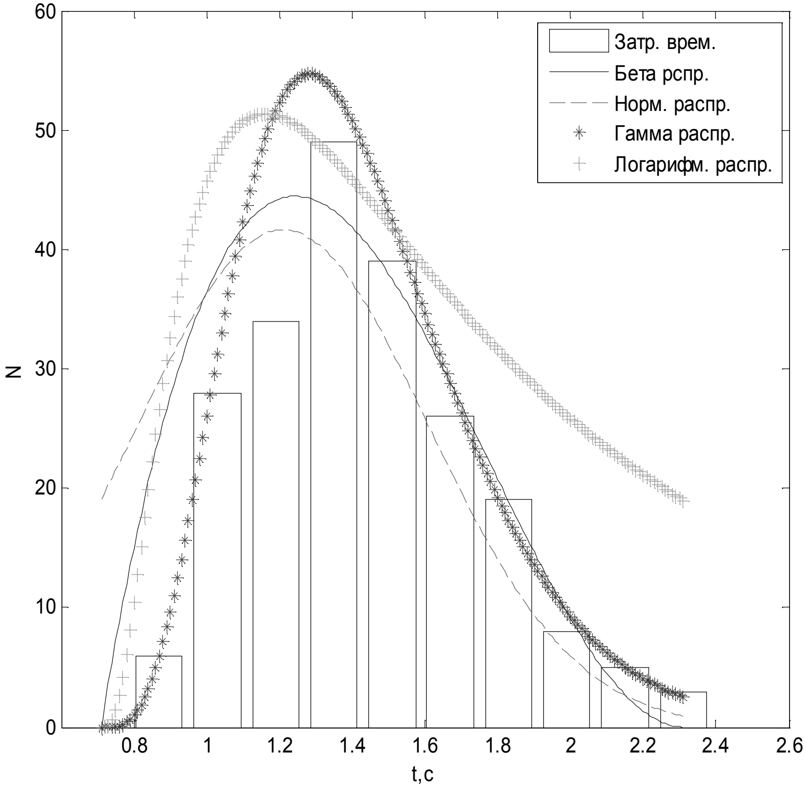


Рис. 1.6. Плотности распределения времени решения оператором задачи по набору цифровых символов

Ко второй группе относятся задачи, в которых оператор при непрерывном наблюдении за состоянием объектов фиксировал изменения их состояния. Задачи отличались количеством одновременно появляющихся изменений и качеством фона (постоянный или переменный).

Результаты проверки согласия эмпирических и теоретических распределений приведены в табл. 1.3.

На рис.1.7 показана гистограмма и теоретические распределения для данной задачи.

К третьей группе относятся задачи, в которых оператор по заданию экспериментатора определял одну характеристику объекта с использованием уточняющих данных. Задачи отличались способом предъявления информации.

Результаты проверки согласия эмпирических и теоретических распределений для задач этой группы приведены в табл. 1.4, а на рис. 1.8 показаны распределения для данной задачи.

Таблица 1.3

Фиксация изменений состояния объектов

№ п/п	Фон	Количество изменений	Значения χ^2			
			Усечен. норм.	Логарифм. норм.	Бета	Гамма
1	Постоянный	от 1 до 4	13,0	13,0	134	7,7
2	Переменный	от 1 до 4	23,9	15,1	74.5	11,9
3	Переменный	от 1 до 10	5,6	67,7	143	6,4

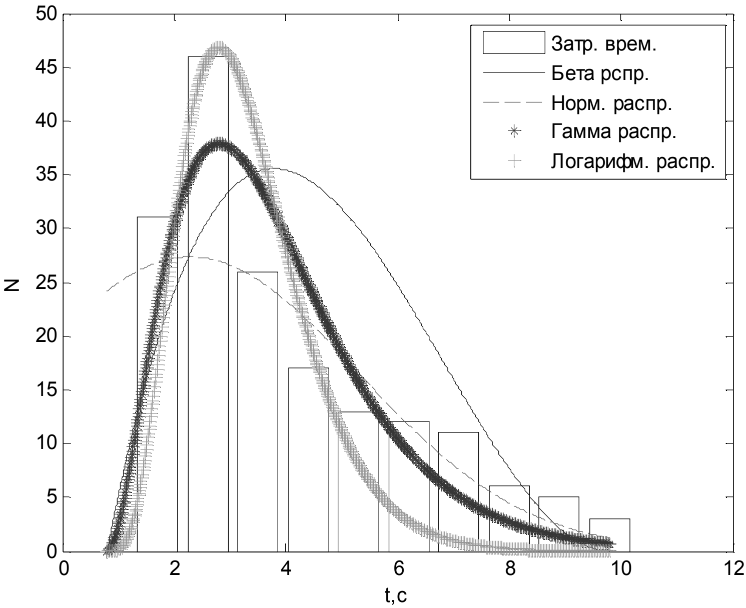


Рис. 1.7. Гистограмма и теоретические распределения времени решения для задачи №1

Таблица 1.4

Определение характеристик объектов с использованием уточняющих данных

№ п/п	Способ предъявления информации	Значения χ^2			
		Усечен. норм.	Логарифм. норм.	Бета	Гамма
1	С выделением строки	9,7	68,3	7,5	38
2	Без выделения	7,9	325	8,0	89
3	Только 1 строка	5,6	16,5	9,5	414

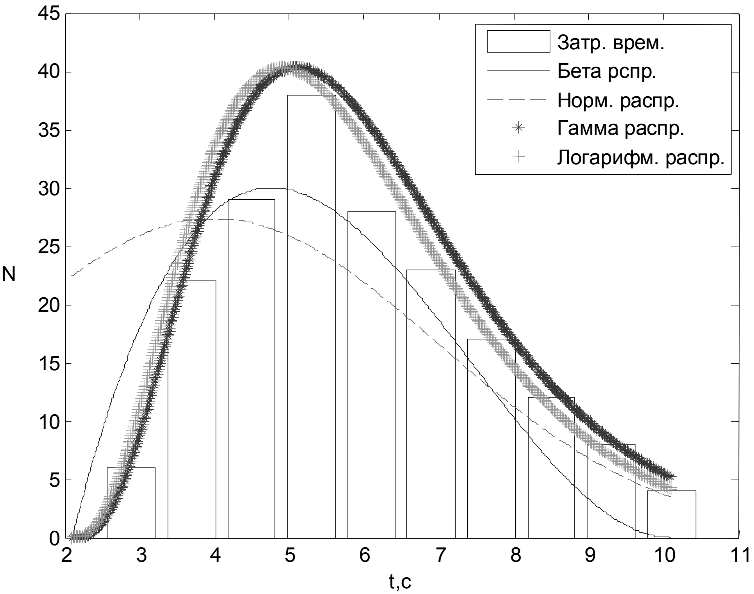


Рис. 1.8. Определение характеристик объектов с использованием уточняющей информации

К четвертой группе относятся задачи, в которых оператор обнаруживал новые объекты на индикаторном устройстве или считывал, либо определял заданную характеристику объекта. Задачи отличались, помимо этого, типом информации об объекте и использованием уточняющих данных.

Результаты исследования приведены в табл. 1.5 и на рис. 1.9.

Таблица 1.5

Определение характеристик цели с использованием БЭ

№ п/п	Тип формуляра	Использование ТХЦ	Тип задачи	Значения χ^2			
				Усеч. норм.	Лог. норм.	Бета	Гамма
1	Короткий	Нет	Обнаруж.	54,3	13,0	68,6	12,9
2	Полный	Нет	Обнаруж.	44,1	8,2	21,2	20,3
3	Короткий	Да	Чтение характ.	16,2	97,4	28,9	40,3
4	Короткий	Нет	Чтение характ.	27,9	19,1	185	9,3
5	Короткий	Нет	Опред. характ.	27,7	4,5	29,0	10,7
6	Полный	Нет	Опред. характ.	14,7	3,2	24,3	2,8

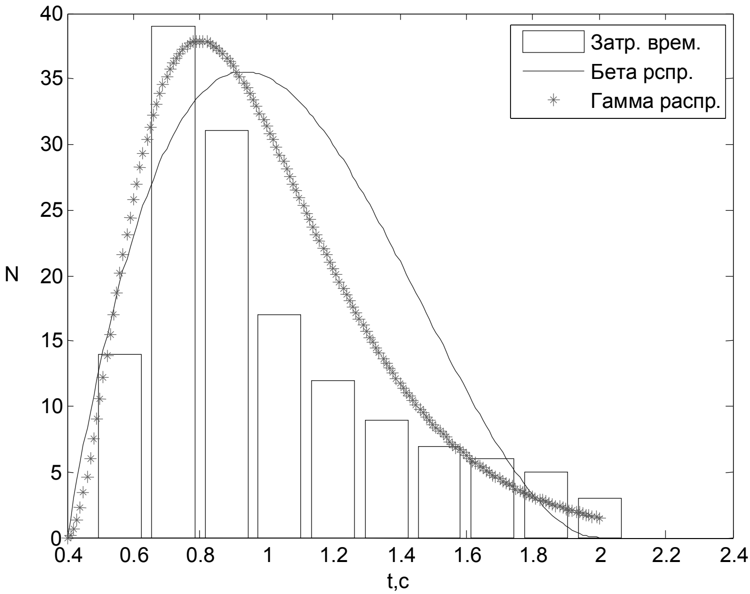


Рис. 1.9. Распределение затрат времени
на обнаружение новых объектов

ГЛАВА 2. ОСНОВЫ ТЕОРИИ ГРАФОВ

Следующим разделом математики, знание которого необходимо специалисту, работающему в области эргономики, несомненно, является теория графов. Данный математический аппарат позволяет решать целый ряд задач, связанных с описанием некоторой совокупности объектов.

2.1. Основные определения графов

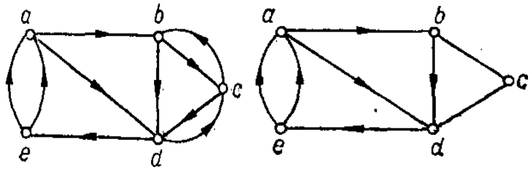
Многие задачи сводятся к рассмотрению совокупности объектов (свойств, операций), существенные свойства которых описываются связями между ними. В подобных случаях удобно рассматриваемые объекты изображать точками, называемыми вершинами, а связи между ними – линиями (произвольной конфигурации), называемыми ребрами.

Множество вершин V, связи между которыми определены множеством ребер E называют графом и обозначают $G = (V, E)$. Примером классической постановки задачи, решаемой с помощью методов графов, является задача, сформулированная Леонардом Эйлером в 1736 г. Это задача о кенигсбергских мостах (рис. 2.1, а): можно ли совершить прогулку таким образом, чтобы выйдя из любого места города, вернуться в него, пройдя в точности один раз по каждому мосту? Так как связи между частями $a, \mathrm{b}, \mathrm{c}, \mathrm{d}$ осуществляются только через семь мостов, то каждый из них изображается ребром, соединяющим соответствующие вершины (рис. 3.1, б).

а б

Рис. 2.1. Задача Эйлерао про кенигсбергские мосты:
а – план города; б - граф

Для указания направления связи между вершинами графа соответствующее ребро отмечается стрелкой. Ориентированное таким образом ребро называют дугой, а граф с ориентированными ребрами – ориентированным графом или короче орграфом (рис. 2.2, а).



а б

Рис. 2.2. Ориентированный (а) и смешанный (б) графы

Если пара вершин соединяется двумя или большим числом дуг, то такие дуги называют параллельными. При этом две дуги, одинаково направленные по отношению к данной вершине, называют строго параллельными, а различно направленные – нестрого параллельными. Нестрого параллельные дуги, отображающие ориентацию связи в обоих направлениях, эквивалентны неориентированной связи и могут быть заменены ребром.

Произведя такую замену в орграфе, придем к смешанному графу, который содержит ребра и дуги (рис. 2.2, б). А любой неориентированный или смешанный граф можно преобразовать в ориентированный заменой каждого ребра парой нестрого параллельных дуг.

Дальнейшее обобщение отображения связей между объектами с помощью графов состоит в приписывании ребрам и дугам некоторых количественных значений, качественных признаков или характерных свойств, называемых весами. Граф, у которого задан вес для каждого ребра или дуги, называется взвешенным. В простейшем случае это может быть порядковая нумерация ребер и дуг, указывающая на очередность при их рассмотрении. Вес ребра или дуги может означать длину (пути сообщения), пропускную способность (трубопровода), количество набранных очков (турниры), количество рядов движения (автомобильные дороги), характер отношении между людьми (начальник, подчиненный) и т. п.

Вес можно приписывать не только ребрам и дугам, но и вершинам. Например, вершины, соответствующие населенным пунктам на карте автомобильных дорог, могут характеризоваться количеством мест в кемпингах, пропускной способностью станций техобслуживания. Вообще, вес вершины может означать любую характеристику соответствующего ей объекта (цвет изображаемого вершиной предмета, возраст человека и т.п.).

2.2. Типы конечных графов

Если множество вершин графа конечно, то он называется конечным графом. Конечный граф $G=(V,E)$, содержащий ровно p вершин и q ребер, называется (p,q)-графом.

Пусть $V=\{v_1,v_2,...,v_p\}$ и $E=\{e_1,e_2,...,e_q\}$ - соответственно, множества вершин и ребер (p,q)- графа. Каждое ребро $e_k \in E$ соединяет пару вершин v_i, $v_j \in V$, являющихся его концами (граничными вершинами). Для ориентированного ребра (дуги) различают начальную вершину, из которой дуга исходит, и конечную вершину, в которую дуга заходит.

Ребро, граничными вершинами которого является одна и та же вершина, называется петлей.

Ребра с одинаковыми граничными вершинами являются параллельными и называются кратными.

В общем случае граф может содержать и изолированные вершины, которые не являются концами ребер и не связаны ни между собой, ни с другими вершинами.

Число ребер, связанных с вершиной v_i (петля учитывается дважды), называют степенью вершины и обозначают через $\delta(v_i)$.

Степень изолированной вершины равна нулю. Вершина степени единицы называется концевой или висячей вершиной.

Легко показать, что в любом графе сумма степеней всех вершин равна удвоенному числу ребер, а число вершин нечетной степени всегда четно.

В орграфе различают положительные $\delta^+(v_i)$ и отрицательные $\delta^-(v_i)$ степени вершин, которые равны соответственно числу исходящих из v_i и заходящих в v_i дуг. Очевидно, суммы положительных и отрицательных степеней всех вершин орграфа равны между собой и равны числу всех дуг.

Пример 2.1. Для конечного (5, 6) – графа на рис. 2.3, а множество вершин $V = \{v_1, v_2, v_3, v_4, v_5\}$; множество ребер $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$; ребра e_2 и e_3 параллельны; ребро e_6 является петлей; v_4 – изолированная вершина и ее степень равна 0 ($\delta(v_4) = 0$); v_1 – висячая вершина и ее степень равна 1 ($\delta(v_1) = 1$); остальные вершины графа имеют такие степени: $\delta(v_2) = 4$, $\delta(v_3) = 3$, $\delta(v_5) = 4$.

Для вершины d орграфа на рис. 2.2, а имеем положительную ($\delta^+(d) = 2$) и отрицательную ($\delta^-(d) = 3$) степени.

Простой или обыкновенный граф – это граф без петель и кратных ребер.

Полный граф – это простой граф, в котором любые две вершины соединены ребром.

Мультиграф – это граф, не содержащий петель, но имеющий кратные ребра.

Псевдограф – это граф, допускающий петли и кратные ребра (наиболее общий случай графа).

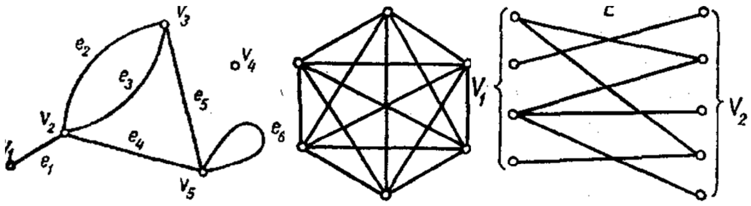
Пустой или нуль-граф – это граф, не имеющий ребер ($E = \varnothing$, все его вершины изолированы).

Если множество вершин V простого графа допускает разбиение на два непересекающихся подмножества V_1 и V_2 ($V_1 \cap V_2 = \varnothing$) такое,что не существует ребер, соединяющих вершины одного и того же подмножества, то он называется двудольным или биграфом.

Ориентированный граф считается простым, если он не имеет строго параллельных дуг и петель.

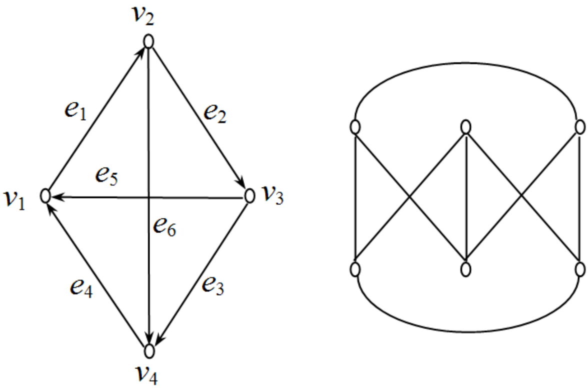
Граф, степени всех вершин которого одинаковы и равны r , называется однородным (регулярным) r -ой степени. Полный граф с n вершинами всегда является однородным графом степени $n - 1$, а пустой граф – однородный граф степени 0. Граф третьей степени называют кубическим графом. Он обладает рядом интересных свойств и, в частности, всегда имеет четное число вершин.

Пример 2.2. На рис. 2.3, а показан псевдограф с петлей e_6 и кратными ребрами e_2 и e_3 . Граф на рис. 2.2, а – это мультиграф с тремя парами кратных ребер. Биграф с двумя непересекающимися подмножествами вершин V_1 и V_2 , показанный на рис. 2.3, в – это простой граф. На рис. 2.3, б показан полный граф степени 5, а на рис. 2.4 – полный (ориентированный) и неполный кубические графы. Орграф, показанный на рис. 2.2, а не является простым.



а б в

Рис. 2.3. Типы графов: а – псевдограф; б – полный граф; в – двудольный граф (биграф)



а б

Рис. 2.4. Кубические графы: а – полный, б – неполный

2.3. Смежность и инцидентность

Две вершины v_i и $v_j \in V$ графа $G=(V,E)$ называются смежными, если они являются граничными вершинами ребра $e_k \in E$.

Отношение смежности на множестве вершин графа можно определить, представив каждое ребро как пару смежных вершин, т.е. $e_k=(v_i,v_j)$, $k=1,2,...,q$. Для неориентированных графов такие пары неупорядочены, т. е. $e_k=(v_i,v_j)=(v_j,v_i)$, а для орграфов − упорядочены, причем v_i *и* v_j, означают соответственно начальную и конечную вершины дуги e_k. Петля при вершине v_i представляется парой (v_i,v_j). Множество вершин V вместе с определенным на нем отношением смежности полностью определяет граф.

Граф можно представить матрицей смежности. Строки и столбцы этой квадратной матрицы соответствуют вершинам графа, а ее (ij) −элемент равен числу кратных ребер, связывающих вершины v_i *и* v_j (или направленных от вершины v_i к вершине v_j для орграфа).

Матрица смежности неориентированного графа всегда симметрична, а орграфа – несимметрична. Неориентированным ребрам соответствуют пары ненулевых элементов, симметричных относительно главной диагонали матрицы, дугам − ненулевые элементы матрицы, а петлям – ненулевые элементы главной диагонали. В столбцах и строках, соответствующих изолированным вершинам, все элементы равны нулю. Элементы матрицы простого графа всегда равны 0 или 1, причем все элементы главной диагонали нулевые.

Пример 2.3. Графы, приведенные на рис. 2.2, а и 2.3, а, описываются матрицами смежности V_1 и V_2:

$$V_1 = \begin{array}{|c|c|c|c|c|l}
\multicolumn{1}{c}{a} & \multicolumn{1}{c}{b} & \multicolumn{1}{c}{c} & \multicolumn{1}{c}{d} & \multicolumn{1}{c}{e} & \\ \hline
0 & 1 & 0 & 1 & 0 & a \\ \hline
0 & 0 & 1 & 1 & 0 & b \\ \hline
0 & 1 & 0 & 1 & 0 & c \\ \hline
0 & 0 & 1 & 0 & 1 & d \\ \hline
2 & 0 & 0 & 0 & 0 & e \\ \hline
\end{array}$$

$$V_2 = \begin{array}{|c|c|c|c|c|l}
\multicolumn{1}{c}{v_1} & \multicolumn{1}{c}{v_2} & \multicolumn{1}{c}{v_3} & \multicolumn{1}{c}{v_4} & \multicolumn{1}{c}{v_5} & \\ \hline
0 & 1 & 0 & 0 & 0 & v_1 \\ \hline
1 & 0 & 2 & 0 & 1 & v_2 \\ \hline
0 & 2 & 0 & 0 & 1 & v_3 \\ \hline
0 & 0 & 0 & 0 & 0 & v_4 \\ \hline
0 & 1 & 1 & 0 & 1 & v_5 \\ \hline
\end{array}$$

Для взвешенного графа, не содержащего кратных ребер, можно обобщить матрицу смежности так, что каждый ее ненулевой элемент равняется весу соответствующего ребра или дуги. Обратно, любая квадратная матрица n -го порядка может быть представлена орграфом с n вершинами, дуги которого соединяют смежные вершины и имеют веса, равные соответствующим элементам матрицы. Если матрица симметрична, то она представима неориентированным графом.

Если вершина v_i является концом ребра e_k, то говорят, что они инцидентны: вершина v_i инцидентна ребру e_k и ребро e_k инцидентно вершине v_i.

В то время как смежность представляет собой отношение между однородными объектами (вершинами), инцидентность – это отношение между разнородными объектами (вершинами и ребрами). При рассмотрении орграфов различают положительную инцидентность (дуга исходит из вершины) и отрицательную инцидентность (дуга заходит в вершину).

Для (p,q)-графа можно построить матрицу инцидентности размера $p \times q$, строки которой соответствуют вершинам, а столбцы – ребрам. Для неориентированного графа элементы этой матрицы определяются по следующему правилу: ij-й элемент равен 1, если вершина v_i инцидентна ребру e_j, и равен нулю, если v_i, и e_j не инцидентны. В случае орграфа ненулевой ij – элемент равен 1 если v_i начальная вершина дуги e_j, и равен -1, если v_i - конечная вершина дуги.

Пример 2.4. Матрицы инцидентности графов, приведенных на рис. 2.3 а и 2.4, а , сотвественно имеют следующий вид:

$$A_1 = \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \left|\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right| \end{array} \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array}$$

$$A_2 = \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \left|\begin{array}{cccccc} 1 & 0 & 0 & -1 & -1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & -1 \end{array}\right| \end{array} \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array}$$

Каждый столбец матрицы инцидентности содержит обязательно два единичных элемента (для орграфа эти элементы всегда имеют различные знаки и равны соответственно *1* и *-1*). Количество единиц в строке равно степени соответствующей вершины (для орграфа количество положительных единиц определяет положительную степень, а количество отрицательных единиц – отрицательную степень). Нулевая строка соответствует изолированной вершине, а нулевой столбец – петле, причем нулевой столбец матрицы инцидентности лишь указывает на наличие петли, но не содержит сведений о том, с какой вершиной эта петля связана (т.е. матрица инцидентности неоднозначно определяет граф).

Графы, для которых сохраняется отношение инцидентности, называются изоморфными.

Пример 2.5. Графы, изображенные на рис. 2.5 имеют такую же матрицу инцидентности (A_2), как и граф рис. 2.4, а, однако с геометрической точки зрения они совершенно различны, хотя по существу различаются лишь начертанием, а отношения инцидентности (при соответствующем обозна-чении вершин и ребер) одинаковы.

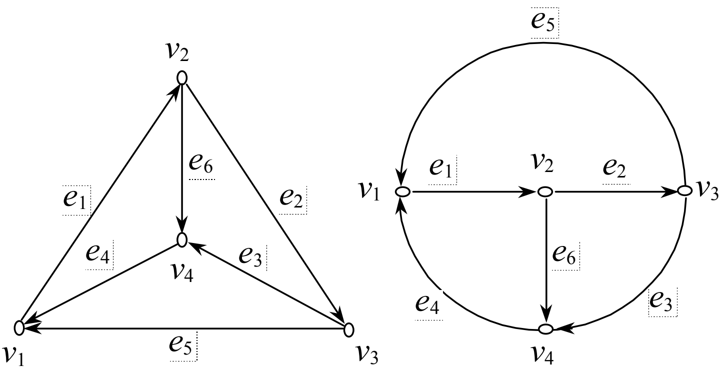


Рис. 2.5. Изоморфные графы

Матрица инцидентности определяет граф без петель с точностью до изоморфизма. Обычно на ее основе можно изобразить различные в геометрическом отношении, но изоморфные между собой графы, каждый из которых называют геометрической реализацией.

Если существенные свойства графа не связаны со способом его изображения на плоскости или нумерацией вершин и ребер, то изоморфные графы, как правило, не различают между собой.

2.4. Маршруты и подграфы

Маршрут длины m на графе – это последовательность m ребер графа (не обязательно различных), таких, что граничные вершины двух соседних ребер совпадают. Маршрут проходит и через все вершины, инцидентные входящим в него ребрам.

Замкнутый маршрут – это маршрут, который приводит в ту же вершину, из которой он начался. Маршрут, все ребра которого различны, называется цепью, а маршрут, для которого различны все вершины, называется простой цепью. Замкнутая цепь называется циклом, а простая замкнутая цепь – простым циклом.

Ориентированные маршруты на орграфе определяются аналогично с той разницей, что начальная вершина каждой последующей дуги маршрута должна совпадать с конечной вершиной предыдущей дуги. Иначе говоря, движение по маршруту допускается только в направлениях, указанных стрелками.

Путь – это маршрут, не содержащий повторяющихся дуг. Простой путь– это путь, не содержащий повторяющихся вершин (за исключением, может быть, начала и конца маршрута). Замкнутый путь называется контуром, а простой замкнутый путь – простым контуром. Граф (орграф) называется циклическим(контурным), если он содержит хотя бы один цикл (контур), в противном случае он называется ациклическим (бесконтурным).

Пример 2.6. Рассмотрим граф, показанный на рис. 2.3, а. На этом графе: $(e_1, e_2, e_3, e_4, e_5, e_6)$ – маршрут, проходящий через последовательность вершин $(v_1, v_2, v_3, v_4, v_5, v_6)$ и соединяющий вершины v_1 и v_5; (e_5, e_6, e_4, e_4) - маршрут, проходящий через последовательность вершин $(v_3, v_5, v_5, v_2, v_5)$, соединяя v_3 и v_5; $(e_1, e_3, e_5, e_4, e_1)$ - замкнутый маршрут; (e_2, e_5, e_6) - цепь; (e_1, e_2, e_5) - простая цепь; (e_2, e_3, e_4, e_5) – цикл; (e_2, e_4, e_5) - простой цикл.

На орграфе (рис. 2.5) маршрут (e_1, e_2, e_5) – простой путь, являющийся контуром, а маршрут (e_1, e_2, e_3) – простой неконтурный путь.

Цикл, который содержит все ребра графа, называется эйлеровым циклом (задача о кенигсбергских мостах сводится к выяснению существования такого цикла), а граф, в котором имеется такой цикл, называется эйлеровым графом. Простой цикл, который проходит через все вершины графа, называют гамильтоновым. Если критерий существования эйлерового цикла очень прост (необходимо, чтобы степени всех вершин были четными), то для гамильтоновых циклов никакого общего правила не найдено.

Понятие цепи применимо и к ориентированным графам. При этом направления дуг не учитываются, т.е. по существу вместо орграфа рассматривают неориентированный, соотнесенный ему граф.

Граф $G' = (V', E')$ является частью графа $G' = (V', E')$, если $V' \subset V$ и $E' \subset E$, т.е. граф содержит все вершины и ребра любой его части.

Подграф – это часть графа, которая, наряду с некоторым подмножеством ребер графа, содержит и все инцидентные им вершины.

Суграф – это часть графа, которая наряду с некоторым подмножеством ребер графа, содержит все вершины графа $(V' = V, E' \subset E)$.

Исходный граф по отношению к его подграфу называют *надграфом*, а по отношению к суграфу – сверхграфом. Совокупность всех ребер графа, не принадлежащих его подграфу (вместе с инцидентными вершинами), образует дополнение подграфа. Говорят, что подграфы одного графа $G' = (V', E)$ и $G'' = (V'', E'')$ разделены ребрами, если они не имеют общих ребер ($E' \cap E'' = \varnothing$) и разделены вершинами, если у них нет общих вершин ($V' \cap V'' = \varnothing$).

Пример 2.7. На рис. 2.6 показаны различные части кубического орграфа G:

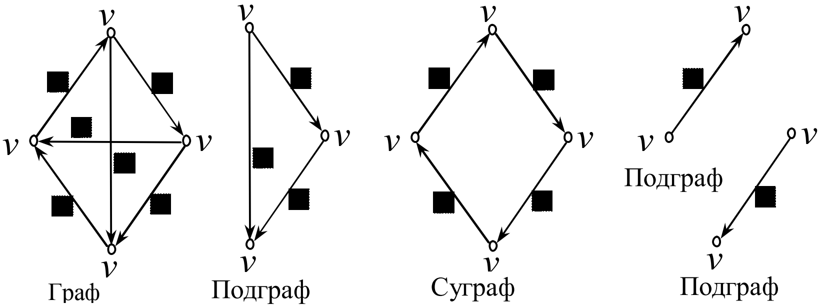


Рис. 2.6. Подграфы и суграфы

2.5. Связность и разделимость

Две вершины графа называют связанными, если существует маршрут, соединяющий эти вершины. Граф, любая пара вершин которого связана, называют связным графом.

В связном графе между любыми двумя вершинами существует простая цепь, так как из связывающего их маршрута всегда можно удалить циклический участок, проходящий через некоторую вершину более одного раза.

Если граф не связный, то множество его вершин можно единственным образом разделить на непересекающиеся подмножества, каждое из которых содержит все связанные между собой вершины и вместе с инцидентными им ребрами образует связный подграф. Таким образом, несвязный граф представляет собой совокупность отдельных частей (подграфов), называемых компонентами (рис. 2.7).

Часто отношение связности усложняется дополнительными условиями. Граф называют циклически связным, если любые две различные вершины содержатся в цикле.

Граф называют k-связным, если любая пара различных вершин связана, по крайней мере k цепями, которые не имеют общих вершин (кроме начальной и конечной).

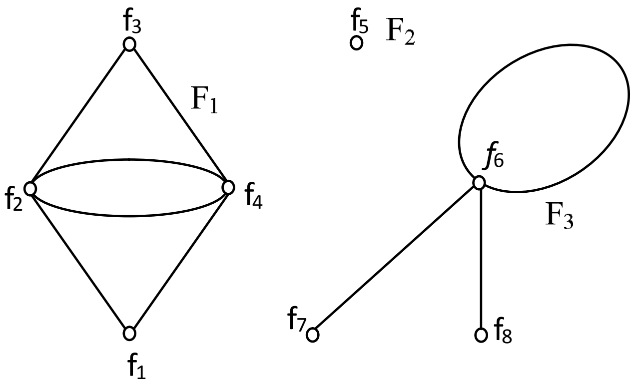


Рис. 2.7. Несвязный граф

Связность ориентированных графов определяется так же, как и для неориентированных (без учета направлений дуг). Специфичным для орграфа (или смешанного графа) является понятие сильной связности. Орграф называют сильно связным, если для любой пары его вершин v_i, v_j существует путь из v_i в v_j и из v_j в v_i.

Пример 2.7. На рис. 2.4, б показан связный граф. Граф F на рис. 2.7 – несвязный, состоящий из трех компонент F_1, F_2, F_3 (изолированная вершина считается компонентой). Компонента F_1 графа F циклически связна (односвязная), а связная компонента F_3 циклически связной не является, так как вершины f_7 и f_8 не содержится ни в каком цикле с другими вершинами.

Орграф G на рис. 2.6 - сильно связный. Также сильно связным всегда должен быть граф, представляющий план города с односторонним движением по некоторым улицам. В противном случае нашлись бы вершины (площади и перекрестки), между которыми нельзя было бы проехать по городу без нарушения правил движения.

Связный граф может быть разделен на несвязные подграфы удалением из него некоторых вершин и ребер (при удалении вершин исключаются и все инцидентные им ребра, а при удалении ребер вершины сохраняются).

Если существует такая вершина, удаление которой превращает связный граф (или компоненту несвязного графа) в несвязный, то она называется точкой сочленения. Ребро с такими же свойствами называется мостом.

При наличии моста в графе имеется, по крайней мере, две точки сочленения.

Граф называется неразделимым, если он связный и не имеет точек сочленения. Граф, имеющий хотя бы одну точку сочленения, является разделимым и называется сепарабельным. Он разбивается на блоки, каждый из которых представляет собой максимальный неразделимый подграф. Каждое ребро графа, как и каждая вершина (за исключением точек сочленения), принадлежат только одному из его блоков. Более того, только одному блоку принадлежит и каждый простой цикл. Значит совокупность блоков графа представляет собой разбиение множеств ребер и простых циклов на непересекающиеся подмножества.

Пример 2.8. Граф G на рис.2.6 неразделим. Связный граф, представленный на рис. 2.8, имеет две точки сочленения – v_1 и v_4, однако его ребро (v_1, v_4) мостом не является. Граф B на рис. 2.9 имеет две точки сочленения – v_4 и v_5, причем ребро (v_4, v_5), соединяющее эти точки, является мостом, который разбивает данный граф на три блока (блоки B_1, B_2 и B_3 на рис. 2.9). Каждый из этих блоков представляет собой неразделимый подграф.

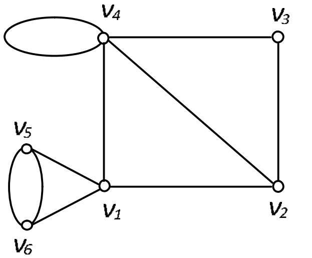


Рис. 2.8. Разделимый граф

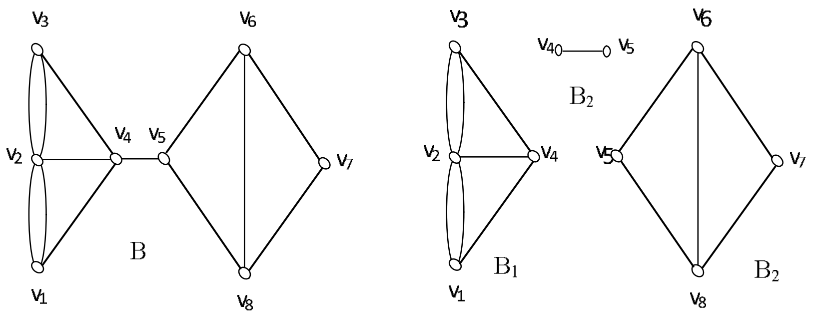


Рис. 2.9. Блоки разделимого графа B с мостом

В ряде приложений теории графов блоки можно рассматривать как компоненты. Это обычно допустимо, когда связи блоков посредством точки сочленения несущественны или когда существенные свойства графа связаны только с его простыми циклами (контурами). В таких случаях можно рассматривать несвязный граф как связный разделимый граф, который образуется путем такого объединения компонент, чтобы каждая из них была блоком (это всегда можно сделать, объединив, например, по одной вершине каждого блока в точку сочленения).

2.6. Деревья и лес

Пусть множество V содержит p вершин, которые пронумерованы порядковыми числами от 1 до p, т.е. $V = \{1,\ 2, \ldots, p\}$. Связав эти вершины ребрами так, чтобы отсутствовали циклы, получим некоторый граф, покрывающий данное множество p вершин.

Дерево – это связный ациклический граф. Дерево на множестве p вершин всегда содержит $q = p - 1$ ребро, т. е. минимальное количество ребер, необходимое для того, чтобы граф был связным. Действительно, две вершины связываются одним ребром, и для связи каждой последующей вершины с предыдущими требуется ребро, следовательно, для связи p вершин необходимо и достаточно $p - 1$ ребро. При $p = 2$ дерево единственно и оно состоит из одной ветви. С увеличением p число различных деревьев t_p быстро возрастает и выражается соотношением:

$$t_p = p^{p-2}\,. \tag{2.1}$$

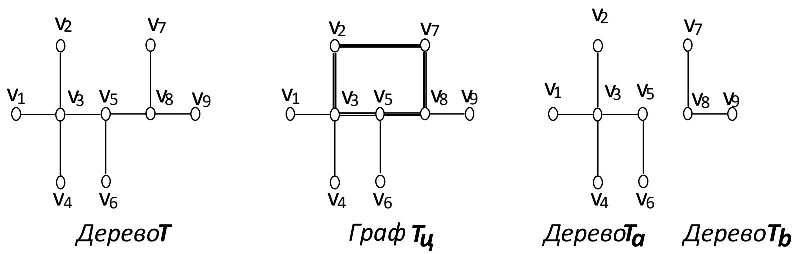
При добавлении в дерево ребра образуется цикл, а при удалении хотя бы одного ребра дерево распадается на компоненты, каждая из которой представляет собой также дерево или изолированную вершину.

Несвязный граф, компоненты которого являются деревьями, называется лесом. Лес из k деревьев, содержащий p вершин, имеет в точности $p - k$ ребер.

Деревья считаются существенно различными, если они не изоморфны. С увеличением числа вершин количество различных деревьев резко возрастает (например, при $p = 20$ их насчитывается около миллиона). Среди различных деревьев выделяются два важных частных случая: последовательное дерево, представляющее собой простую цепь, и звездное дерево, в котором одна из вершин (центр) смежна со всеми остальными вершинами.

Рассматриваются также деревья с ориентированными ребрами (дугами). Ориентированное дерево называется *прадеревом* с корнем v_0, если существует путь между вершиной v_0 и любой другой его вершиной. Ясно, что прадерево имеет единственный корень (вершину, в которую не входит ни одна дуга.

Пример 2.9. Дерево *T* на рис. 2.10, а имеет 9 вершин и 8 ребер. Ввод дополнительного ребра (v_2, v_7) нарушает ацикличность дерева и превращает его в связый граф T_c, в котором цикл $(v_2,\ v_7, v_8, v_5, v_3, v_2)$ выделен двойной линией. Удаление ребра $(v_5,\ v_8)$ превращает дерево T в лес, состоящий из двух деревьев – T_a и T_b. (рис. 2.10, в).



а б в

Рис. 2.10. Дерево Т и варианты его изменения

На рис. 2.11 показаны все существенно различные шестивершинные деревья $(T_1 - T_6)$, среди которых есть последовательное дерево - T_1 и звездное дерево - T_6.

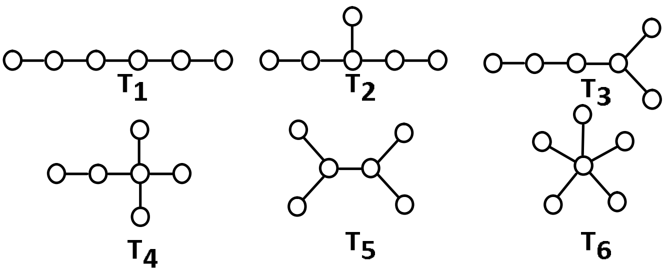


Рис. 2.11. Шестивершинные деревья

На рис. 2.12 показано дерево с ориентированными ребрами, которое, очевидно, является прадерево с корнем v_0.

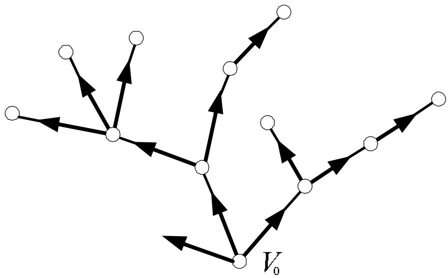


Рис. 2.12. Прадерево с корнем

Деревья играют важную роль в различных прикладных задачах, например, о связи каких-либо объектов минимальным числом каналов с определенными свойствами; об определении координат при моделировании цепей и систем различной физической природы, о моделировании при рассмотрении иерархических систем объектов, структурных формул органических соединений и т. п.

Граф называют плоским (планарным), если существует изоморфный ему граф (геометрическая реализация), который может быть изображен на плоскости без пересечения ребер. Свойства планарности не нарушаются, если некоторое ребро разбить на два введением новой вершины второй степени или заменить два ребра, инцидентные вершине второй степени, одним ребром, удалив эту вершину.

Два графа называют гомеоморфными (изоморфными с точностью до вершин второй степени), если после удаления из них вершин второй степени и объединения инцидентных этим вершинам ребер, они оказываются изоморфными. Очевидно, граф, гомеоморфный плоскому графу, также плоский.

2.7. Практическое использование методов теории графов при моделировании деятельности операторов системы «человек-машина»

В процессе практических исследований часто возникает необходимость разработки моделей деятельности лица, принимающего решения (ЛПР). Так модель деятельности ЛПР по определению положения воздушных объектов и оценке воздушной обстановки (ВО) может быть представлена графом, изображенным на рис. 2.13, 2.14.

В представленных графах вершины соответствуют событиям, например, «информация, представленная на большом экране (БЭ), воспринята», «команды в ЭВМ введены», а ребрам соответствуют вероятности перехода от одного события к другому, а также время, затрачиваемое на такой переход. Содержание вершин графа, и смысл и последовательность переходов между вершинами представлен в табл. 2.1, 2.2.

И так модель деятельности ЛПР формально может быть задана следующим образом:

$$P = \left| p_{ij} \right|, \tag{2.2}$$

$$T = \left| t_{ij} \right|, \tag{2.3}$$

где P – матрица вероятностей переходов между событиями ij;

T – матрица времени, затрачиваемого на работу при переходе от события i к событию j;

p_{ij} – вероятность перехода от события i к событию j;

t_{ij} – время, затрачиваемое на переход от события i к событию j.

На рис. 2.13, 2.14 значения p_{ij} и t_{ij} заданы как w_{ij}, где $w_{ij}\,(p_{ij}, t_{ij})$.

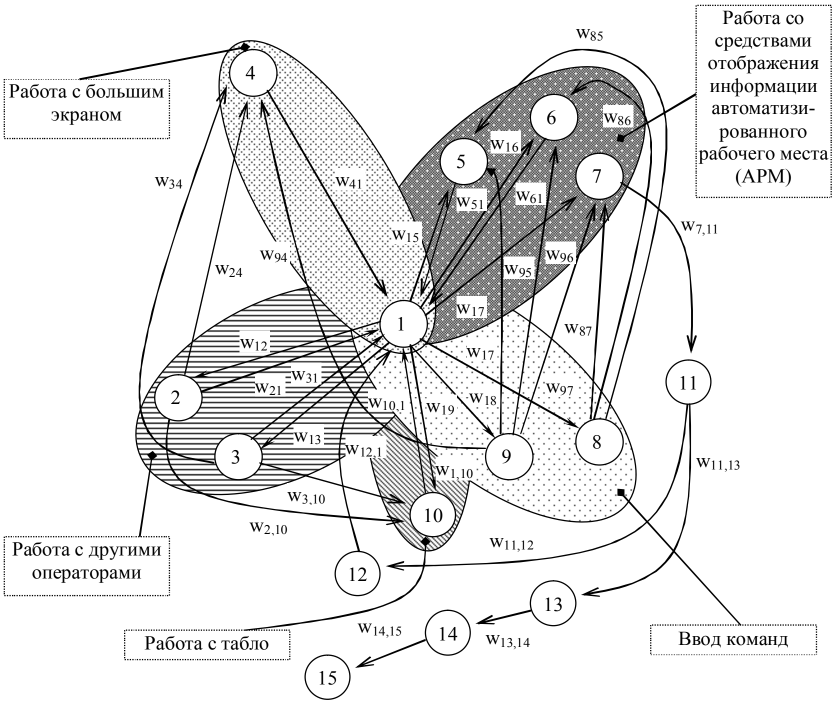


Рис. 2.13. Модель деятельности ЛПР при оценке воздушной обстановки

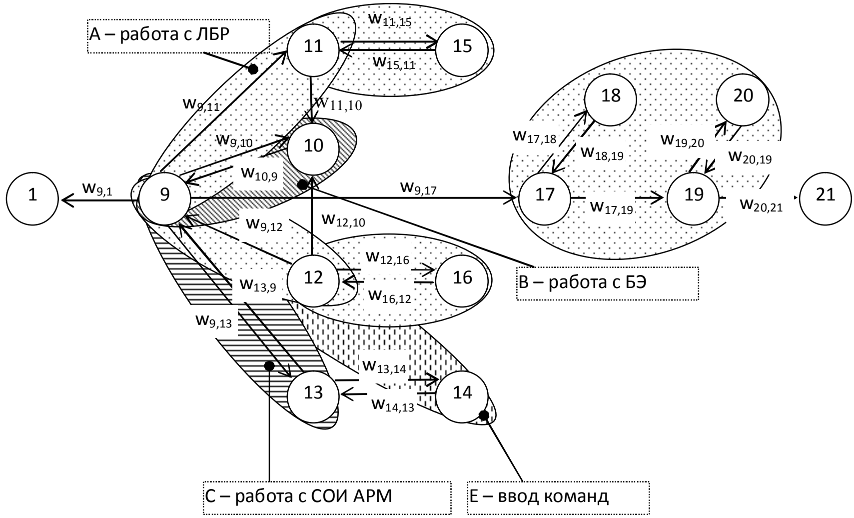


Рис. 2.14. Модель деятельности ЛПР при ликвидации аварийных ситуаций с использование комплекса средств автоматизации

Таблица 2.1

События процесса оценки воздушной обстановки

События	Содержание события
1	Начало работы ЛПР по решению задач оценки ВО
2, 10	Восприятие информации, отображаемой на большом экране БЭ
3,11,4, 12	Поиск дополнительной (уточнение) информации о ВО
5,13	Оценка ВО по информации, представленной на средствах отображения информации(СОИ)(АРМ)
6,14	Ввод команд управления для изменения состава информации, отображаемой
7,15	Работа с другими операторами по получению требуемой информации о ВО (воздушном объекте)
8,16	Работа по уточнению качества источников получения информации о ВО
9	Переход ЛПР к решению задачи ликвидации аварийной ситуции
17,18	Выработка предложений на улучшения качества технической разведки воздушного пространства
19,20	Доклад выводов из оценки ВО
21	Окончание работ

Таблица 2.2

Содержание переходов из состояния в состояние

Переходы	Действия, совершаемые ЛПР при переходе от одного состояния в другое
$w_{12}, w_{9,10}$	Оценка ВО по БЭ
$w_{21}, w_{10,9}$	Поиск дополнительной информации о ВО
$w_{13}, w_{9,11}$ $w_{14}, w_{9,12}$	Обращение к другим операторам для поиска или уточнения информации о ВО (воздушном объекте), либо для изменения состава отображаемой на БЭ информации
$w_{15}, w_{9,13}$	Оценка информации отображаемой на СОИ АРМ
w_{65}, w_{56}, $w_{13,14}, w_{14,13}$	Ввод команд изменения параметров отображения информации на СОИ АРМ (селекция, изменение видов формуляров и др.) и контроль (обрабатывает) результатов вода команд
w_{73}, w_{37}, $w_{15,11}, w_{11,15}$, w_{48}, w_{84}, $w_{12,16}, w_{16,12}$	Поиск и получение дополнительной информации про ВО (ВЦ)
$w_{9,17}$	Информация о ВО ЛПР оценена
$w_{19,21}$	Ввод в управляющий комплекс значения средней степени воздействия по целям, окончание этапа оценки ВО

При рассмотрении модели деятельности ЛПР можно выделить несколько ее фрагментов:

A - работа ЛПР с другими операторами,

B – работа ЛПР с СОИ коллективного пользования,

C – работа ЛПР с СОИ АРМ,

E – действия ЛПР, связанные с вводом команд в ЭВМ,

представленные на рис 2.13, 2.14 соответствующими заштрихованными областями.

Использование теории графов и учет особенностей деятельности оператора позволяет представить модель его деятельности в следующем виде (рис. 2.15), который является определенным отражением моделей, представленных на рис. 2.13, 2.14.

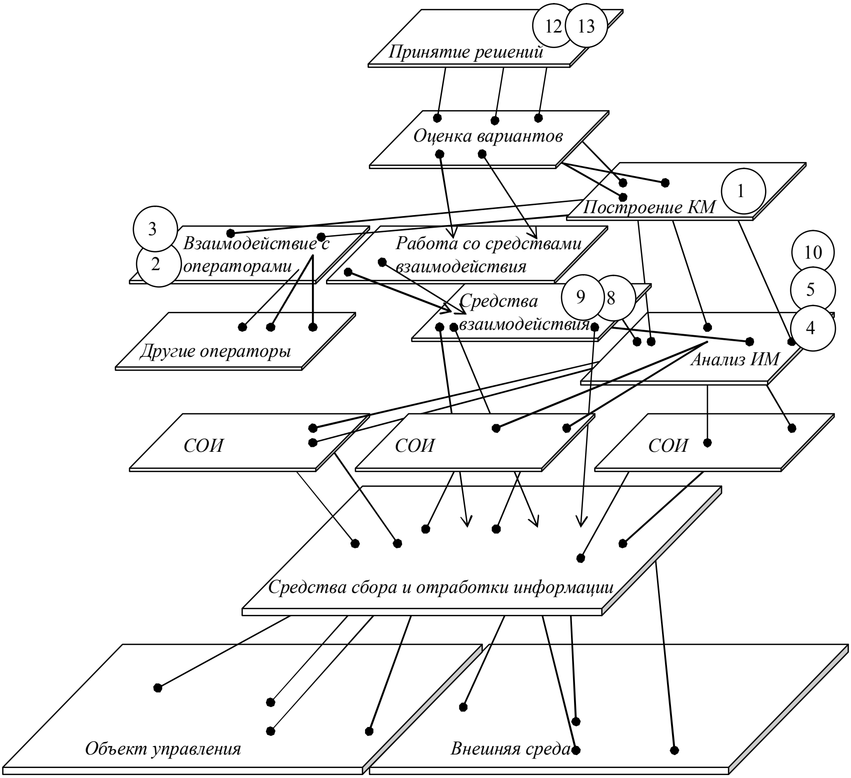


Рис. 2.15. Вариант построения модели деятельности оператора

Использование элементов теории графов в данном примере позволяет исследовать затраты времени оператора на выполнение различного рода операций и проанализировать деятельность оператора в различных условиях.

ГЛАВА 3. ОСНОВЫ ТЕОРИИ МОДЕЛИРОВАНИЯ

Пожалуй, одним из основных процессов, с которым приходится сталкиваться при проведении эргономических исследований, является процесс моделирования тех или иных ситуаций и объектов окружающего мира. Математической основой данного процесса является математическая теория моделирования, основы которой и изложены в данной главе.

3.1. Основные определения

Моделирование – это замещение объекта (процесса, явления) исследования его моделью, отражающей основные свойства объекта. Объект исследования определяется совокупностью параметров и характеристик. Множество параметров системы отражает ее внутреннее содержание – структуру и принципы функционирования. Характеристики системы – это ее внешние свойства, которые важны при взаимодействии с другими системами. Характеристики системы находятся в функциональной зависимости от ее параметров.

Теория моделирования представляет собой взаимосвязанную совокупность положений, определений, методов и средств создания и изучения моделей. Эти положения, определения, методы и средства, как и сами модели, являются предметом теории моделирования.

Основная задача теории моделирования заключается в создании моделей, которые достаточно точно и полно фиксируют интересующие свойства оригиналов, проще или быстрее поддаются исследованию и допускают перенесение его результатов на оригиналы.

Основная задача математического моделирования – выделение законов в природе, обществе и технике и запись их на языке математики.

Модель объекта – это физическая или абстрактная система, адекватно представляющая объект исследования. В теории моделирования используются преимущественно абстрактные модели – описания объекта исследования на некотором языке. Абстрактность модели проявляется в том, что компонентами модели являются не физические элементы, а понятия, в качестве которых наиболее широко используются математические. Абстрактная модель, представленная на языке математических отношений, называется математической моделью. Математическая модель имеет форму функциональной зависимости $Y = F(X)$, где $Y = \{y_1, y_2, ..., y_m\}$ и $X = \{x_1, x_2, ..., x_n\}$ – соответственно характеристики и параметры моделируемой системы, а F – функция, воспроизводимая моделью. Построение модели сводится к выявлению функции F и представлению ее в форме, пригодной для вычисления значений $Y = F(X)$. Модель позволяет оценивать характеристики Y для заданных параметров X и выбирать значения параметров, обеспечивающие требуемые характеристики, с использованием процедур оптимизации.

Модель сложного объекта (процесса, системы) не может быть простой. Из чего следует, что процесс использования математических моделей реальных систем является итерационным процессом, когда последовательно уточняется (дорабатывается) математическая модель и методы решения стоящих задач.

Модель может воспроизводить полную совокупность свойств системы либо их подмножество. Состав свойств устанавливается в зависимости от цели исследований и при построении модели определяет:

– состав воспроизводимых характеристик $Y = \{y_1, y_2, ..., y_m\}$;

– состав параметров $X = \{x_1, x_2, ..., x_n\}$, изменение которых должно влиять на характеристики Y;

– область изменения параметров $x_i \in x_j^*$, $j = 1, 2, ..., n$ - область определения модели;

– точность – предельная допустимая погрешность оценки характеристик Y на основе модели.

Состав характеристик Y определяется в зависимости от исследуемых свойств системы – производительности, надежности, стоимости и других свойств и должен гарантировать полноту отображения этих свойств. Состав параметров X должен охватывать все существенные аспекты организации системы, изучение влияния которых на качество функционирования составляет цель моделирования.

Область определения модели характеризует диапазон исследуемых вариантов организации системы. Чем шире состав характеристик и параметров, а также область определения модели, тем универсальнее модель в отношении задач, которые можно решать с ее использованием.

Допустимые погрешности оценки характеристик и точность задания параметров определяют требования к точности модели. Так, если изменения характеристик в пределах 10 % несущественны для выбора варианта системы, то точность определения характеристик должна составлять $\pm 5\%$. В большинстве случаев параметры, и в первую очередь параметры рабочей нагрузки, могут быть заданы лишь приближенно, с относительной погрешностью $10-25\%$. В таких случаях нет смысла предъявлять высокие требования к точности воспроизведения моделью характеристик системы и погрешности их оценки на уровне $5-15\%$ вполне приемлемы.

Модель, удовлетворяющая вышеперечисленным требованиям по составу характеристик и параметров и точности воспроизведения характеристик по всей области определения, называется адекватной системе.

Существенное влияние на адекватность оказывает область определения модели. Практически любая модель обеспечивает высокую точность воспроизведения характеристик в пределах малой окрестности точки $X = \{x_1, x_2, ..., x_n\}$. Чем шире область определения модели, тем меньше шансов, что некоторая модель окажется адекватной системе.

Другое свойство модели – сложность. Сложность модели принято характеризовать двумя показателями: размерностью и сложностью вычислений, связанных с определением характеристик. Размерность модели – число величин, представляющих в модели параметры и характеристики. Сложность вычислений, выполняемых при расчете характеристик $Y = F(X)$, оценивается числом операций, приходящихся на одну реализацию оператора F. Обычно сложность вычислений связывается с затратами ресурсов ЭВМ и характеризуется числом процессорных операций и емкостью памяти для хранения информации, относящейся к модели. Сложность вычислений – монотонно возрастающая функция размерности модели. Поэтому более сложной модели присущи одновременно большая размерность и сложность вычислений.

Сложность модели определяется также сложностью моделируемой системы и назначением модели (состав характеристик и параметров, воспроизводимых моделью), размером области определения и точностью модели. Чем сложнее система, то есть чем больше число входящих в нее элементов и процессов, из которых слагается функционирование системы, тем сложнее модель. Увеличение числа воспроизводимых характеристик и параметров, области определения и точности оценки характеристик приводят к увеличению сложности модели.

3.2. Обзор методов моделирования

Модели объектов делятся на два больших класса: материальные (физические) и абстрактные (математические). Среди физических моделей наибольшее распространение получили аналоговые модели.

Классификация математических моделей

Существует несколько схем классификации математических моделей. Все они достаточно условны.

Одна из таких схем приведена на рис. 3.1.

Все математические модели по использованному формальному языку можно разбить на аналитические и имитационные.

Рис. 3.1. Классификация математических моделей

Аналитические – модели, в которых используется стандартный математический язык.

Имитационные – модели, в которых использован специальный язык моделирования или универсальный язык программирования.

Аналитические модели могут быть записаны в виде формул или уравнений. Если какой-либо процесс не может быть описан в виде аналитической модели, его описывают с помощью специального алгоритма или программы. Такая модель является имитационной.

Аналитические модели в свою очередь разбиваются на теоретические и эмпирические модели. Теоретические модели отражают реальные структуры и процессы в исследуемых объектах, то есть, опираются на теорию их работы. Эмпирические модели строятся на основе изучения реакций объекта на изменение условий окружающей среды. При этом теория работы объекта не рассматривается, сам объект представляет собой так называемый «черный ящик», а модель – некоторую интерполяционную зависимость. Эмпирические модели могут быть построены на основе экспериментальных данных. Эти данные получают непосредственно на исследуемых объектах или с помощью их физических моделей.

По форме описания аналитические модели подразделяются на линейные и нелинейные.

Если все входящие в модель величины не зависят от времени, то имеем статическую модель объекта или процесса, в противном случае получаем динамическую модель.

В детерминированных моделях все взаимосвязи, переменные и константы заданы точно, что приводит к однозначному определению результирующей функции. Если часть или все параметры, входящие в модель по своей природе являются случайными величинами или случайными функциями, то модель относят к классу стохастических моделей.

В стохастических моделях задаются законы распределения случайных величин, что приводит к вероятностной оценке результирующей функции.

Если аналитическое исследование может быть доведено до конца, модели называются аналитически разрешимыми. В противном случае говорят о численно разрешимых аналитических моделях.

Создание математической модели преследует две основные цели:

1) формализованное описание структуры и процесса функционирования системы для однозначности их понимания;

2) представление процесса функционирования в виде, допускающем аналитическое исследование системы.

Единой методики построения математических моделей не существует. Это обусловлено большим разнообразием классов систем:

– статические и динамические;

– структурным или программным управлением (в свою очередь, жестким или гибким)

– постоянной и переменной структурой.

По характеру входных воздействий и внутренних состояний системы подразделяются:

– непрерывные и дискретные;

– линейные и нелинейные;

– стационарные и нестационарные;

– детерминированные и стохастические.

3.3. Основные этапы моделирования

При моделировании необходимо создать модель и провести ее исследование, практически всегда на ЭВМ. Моделирование на ЭВМ предполагает выполнение следующих этапов:

1. Формулирование цели моделирования.
2. Разработка концептуальной модели.
3. Подготовка исходных данных.
4. Разработка математической модели.
5. Выбор метода моделирования.
6. Выбор средств моделирования.
7. Разработка программной модели.
8. Проверка адекватности и корректировка модели.
9. Планирование экспериментов.
10. Моделирование на ЭВМ, анализ результатов моделирования.

Рассмотрим представленную последовательность этапов подробнее.

3.3.1. Цель моделирования

Для одной и той же системы S_0 можно построить множество различных моделей $\{S_m\}$.

Эффективность системы должна быть определена до её физической реализации. Если предполагается несколько вариантов системы, то можно выбрать наилучший вариант её построения. Эффективность системы сопоставляется с затратами и достигнутыми характеристиками:

$$E = E\left(Y_0\right);$$

где E − показатель эффективности; Y_0 − множество характеристик.

Однокритериальная оценка ограничивается оценкой эффективности системы по одному частному показателю качества y_{opt}, $y_{opt} \in Y$. По остальным характеристикам накладываются ограничения на их допустимые изменения:

$$E = y_{opt};$$

$$y_{i_{\min}} \le y_i \le y_{i_{\max}}, \quad i = 1, 2, ..., n,$$

где $y_{i_{\min}}, y_{i_{\max}}$ − нижний и верхний пределы i-го частного показателя качества;

n – число учитываемых характеристик.

При многокритериальной оценке часто используется нормированный аддитивный критерий:

$$E = \sum_{i=1}^{n} b_i \gamma\left(y_i\right).$$

Функции $\gamma(y_i)$ подобраны так, чтобы исключить размерность i-й характеристики и обеспечить условие $\gamma(y_i) \in [0,1]$.

Весовые коэффициенты b_i удовлетворяют условию:

$$\sum_{i=1}^{n} b_i = 1; \qquad b_i > 0 .$$

В частном случае, когда $0 \le y_i \le y_{i_{\max}}$, $\gamma(y_i) = y_i / y_{i_{\max}}$, нормированный аддитивный критерий принимает линейную форму.

3.3.2. Создание концептуальной модели

3.3.2.1. Ориентация

При разработке модели выделяются следующие этапы описания: концептуальный, математический, программный.

Концептуальная (содержательная) модель в словесной форме определяет состав и структуру системы, свойства компонентов и причинно-следственные связи между ними. Здесь приводятся сведения и природе и параметрах элементарных явлений исследуемой системы, о виде и степени взаимодействия между ними, о месте и значении каждого явления в общем процессе функционирования системы.

Процесс создания концептуальной модели никогда не будет формализован, поэтому иногда говорят, что моделирование является не только наукой, но и искусством.

3.3.2.2. Стратификация

Стратификация – это выбор уровня детализации модели. Уровни детализации иногда называют стратами. Выбор уровня детализации часто определяется параметрами, допускающими варьирование в процессе моделирования. Такие параметры обеспечивают определение интересующих характеристик. Остальные параметры должны быть, по возможности, исключены из модели.

3.3.2.3. Детализация

Здесь вводится понятие элементарной операции, для которой известны и могут быть получены зависимости выходных параметров от входных. Если при этом получается модель большой размерности, то ее можно преобразовать в иерархическую модульную структуру. Каждый модуль такой структуры состоит из совокупности модулей более низких рангов, в том числе и элементарных операций.

3.3.2.4. Структуризация. Управление.

Связи между компонентами системы могут быть вещественными и информационными. Вещественные отражают возможные пути перемещения продукта преобразования от одного элемента к другому. Информационные связи обеспечивают передачу между компонентами управляющих воздействий и информации о состоянии.

В простых системах информационные связи могут отсутствовать, что соответствует принципу структурного управления. В более сложных системах управление включает указания, какому компоненту, какой исходный объект, когда и откуда взять, какую операцию по преобразованию выполнить и куда передать. Такие системы функционируют в соответствии с программным или алгоритмическим принципом управления. В концептуальной модели должны быть конкретизированы все решающие правила или алгоритмы управления рабочей нагрузкой, компонентами и процессами.

3.3.2.5. Локализация

Внешняя среда представляется в виде генераторов внешних воздействий, включаемых в состав модели в качестве компонентов. Сюда входят генераторы рабочей нагрузки, генераторы управляющих и возмущающих воздействий. Приемники выходных воздействий обычно не включают в модель.

3.3.2.6. Выделение процессов

Функционирование системы заключается в выполнении технологических процессов преобразования вещества, энергии или информации. В сложных системах, как правило, одновременно протекает несколько процессов. Каждый процесс состоит из определенной последовательности отдельных элементарных операций. Часть операций может выполняться параллельно разными активными компонентами системы. Задается технологический процесс одним из видов представления алгоритмов. В системах с программным управлением, обеспечивающим параллельное выполнение нескольких процессов, имеются алгоритмы управления совокупностью параллельно функционирующих процессов.

3.3.3. Подготовка исходных данных

3.3.3.1. Сбор фактических данных

Концептуальная модель определяет совокупность параметров S_0 и внешних воздействий X. Для количественных параметров необходимо определить их конкретные значения, которые будут использованы в виде исходных данных при моделировании. Сбор данных должен учитывать следующее:

– значения параметров могут быть детерминированными и стохастическими;

– не все параметры являются стационарными;

– речь идет о несуществующей системе (проектируемой, модернизируемой).

Ряд параметров являются случайными величинами по своей природе. Для упрощения модели часть из них представляются детерминированными средними значениями. Это допустимо, если величина имеет небольшой разброс, или, когда цель моделирования достигается при использовании средних значений.

Иногда детерминированные параметры представляются случайной величиной. Например, при многократном выполнении программы количество обрабатываемых данных может изменяться. Всю совокупность вариантов данных можно представить случайной величиной с заданным законом распределения вероятностей.

3.3.3.2. Подбор закона распределения

Для случайных параметров выявляется возможность представления их теоретическими законами распределения. Процедура подбора вида закона распределения заключается в следующем.

По совокупности значений параметра строится гистограмма относительных частот – эмпирическая плотность распределения. Гистограмма аппроксимируется плавной кривой, которая сравнивается с кривыми плотности распределения различных теоретических законов распределения. По наилучшему совпадению выбирается один из законов. Далее по эмпирическим значениям вычисляют параметры этого распределения. Проверка совпадения эмпирического и теоретического распределения осуществляется по одному из критериев согласия: Пирсона (χ^2), Колмогорова, Смирнова, Фишера, Стьюдента.

Особую сложность представляет сбор данных по случайным параметрам, зависящим от времени, что характерно для внешних воздействий. Пренебрежение фактами нестационарности параметров существенно влияет на адекватность модели.

3.3.3.3. Аппроксимация функций

Для каждого компонента системы существует функциональная связь между параметрами входных воздействий и выходными характеристиками. Иногда вид функциональной зависимости может быть легко выявлен. Однако для некоторых компонентов удается получить лишь экспериментальные данные по входным параметрам и выходным характеристикам.

В этом случае вводится некоторая гипотеза о характере функциональной зависимости, т.е. аппроксимация ее некоторым математическим уравнением. Поиск математических зависимостей между двумя или более переменными по собранным данным может выполняться с помощью методов регрессионного, корреляционного или дисперсионного анализа.

Вид математического уравнения задает исследователь. Затем методом регрессионного анализа вычисляются коэффициенты уравнения. Приближение кривой к экспериментальным данным оценивается по критерию наименьших квадратов.

Для выяснения того, насколько точно выбранная зависимость согласуется с опытными данными, используется корреляционный анализ.

При проектировании новой системы отсутствует возможность сбора фактических данных. Для таких параметров выдвигаются гипотезы об их возможных значениях.

Этап сбора и обработки данных заканчивается классификацией на внешние и внутренние, постоянные и переменные, непрерывные и дискретные, линейные и нелинейные, стационарные и нестационарные, детерминированные и стохастические. Для переменных количественных параметров определяются границы их изменений, а для дискретных – возможные значения.

3.3.4. Разработка математической модели

Функционирование системы заключается в реализации определенного процесса, заданного целью. При этом в системе одновременно может протекать несколько процессов. Процесс представляет собой определенную последовательность отдельных операций. Часть операций может выполняться параллельно разными активными ресурсами системы. Задается технологический процесс одним из способов представления алгоритмов.

В системах с программным принципом управления, обеспечивающим параллельное выполнение нескольких технологических процессов, имеются алгоритмы управления совокупностью процессов. Они предназначены для разрешения конфликтных ситуаций, возникающих, когда два или более процесса претендуют на один и тот же ресурс. Совокупность алгоритмов управления A_0 совместно с параметрами входных воздействий X_0 и компонентов S_0 описывают динамику функционирования системы. Обычно алгоритмы управления представляются в виде A_m, удобном для моделирования. Данный подход к описанию динамики работы системы наиболее удобен для имитационного моделирования и является естественным способом определения характеристик системы:

$$Y = F(X, S, A, T), \tag{3.1}$$

где F – множество операторов определения выходных характеристик, T – календарное время моделирования.

Для систем со структурным принципом управления получил распространение другой подход к описанию динамики работы системы. Для каждого компонента выбирается определенный параметр S или несколько параметров, значения которых изменяются в ходе функционирования компонента и отражают его состояние в текущий момент времени $z(t)$. Множество таких параметров по всем $n = 1, N$ элементам системы $\{z_n\}$ отражает состояние системы $Z(t)$. Функционирование системы представляется в виде последовательной смены состояний: $Z(t_0)$, $Z(t_1), ..., Z(T)$. Множество $\{Z\}$ возможных состояний системы называют пространством состояний. Текущее состояние системы в момент времени $t(t_0 < t \leq T)$ отражается в виде координаты точки в n–мерном пространстве состояний, а вся реализация процесса функционирования системы за время T – в виде некоторой траектории (3.1).

При заданном начальном состоянии системы $Z^0 = Z(t_0)$ можно определить ее состояние в любой момент t из интервала $[t_0, T]$, если известна зависимость

$$Z(t) = H(X, Z^0, Z, t). \tag{3.2}$$

В этом случае выходные характеристики системы определяются по выражению

$$Y = G(Z, T). \tag{3.3}$$

В выражениях (3.1), (3.3) операторы F и G называют операторами выходов, а в (3.2) оператор H – оператором переходов.

Для перевода обобщенной модели в форме (3.1) или в форме (1), (3) в конструктивную необходимо конкретизировать свойства множеств переменных и операторов.

3.3.4.1. Математические модели аналитического типа

Простейшие аналитические модели могут быть заданы явно в виде функции одной или нескольких переменных.

Обычно в виде функций задаются общие законы природы или общие закономерности, полученные в результате интегрирования дифференциальных уравнений. Примером такой модели может служить знаменитая формула К.Э. Циолковского:

$$\Delta v_{\text{ла}} = v \ln \frac{M_0}{M_к},$$

определяющая приращение скорости ракеты $\Delta v_{\text{ла}}$ при импульсном сжигании топлива через скорость истечения рабочего тела v и отношение начальной M_0 и конечной $M_к$ масс ракеты.

Модель, заданная в явном виде, дает исчерпывающее описание исследуемого объекта. Она позволяет построить зависимость его характеристик от управляющих факторов, взять производные и найти экстремумы модели, определить характеристики модели в окрестности экстремумов и т.д.

Очень удобна графическая интерпретация таких моделей. Однако модели в аналитическом виде могут быть разработаны только для ограниченного класса объектов, процессов или явлений.

3.3.4.2. Линейные математические модели

Наиболее простыми являются так называемые линейные детерминированные модели. Они задаются в виде линейной формы управляющих переменных x:

$$W = a_0 + a_1 x_1 + ... + a_k x_k,$$

при линейных ограничениях вида

$$\begin{aligned} &b_{1j}x_1 + b_{2j}x_2 + ... + b_{kj}x_k \ge b_j,\ j = 1, 2, ..., q_1; \\ &c_{1j}x_1 + c_{2j}x_2 + ... + c_{kj}x_k = c_j,\ j = 1, 2, ..., q_2; \\ &d_{1j}x_1 + d_{2j}x_2 + ... + d_{kj}x_k \le d_j,\ j = 1, 2, ..., q_3. \end{aligned} \tag{3.4}$$

Общее число ограничений $m = q_1 + q_2 + q_3$ может превосходить число переменных ($m > k$). Кроме того, обычно вводится условие положительности переменных ($x_i \ge 0$).

Поверхность отклика для линейной модели представляет собой гиперплоскость.

Пример 3. Рассмотрим линейную модель двух переменных следующего вида:

$$W = -2x_1 - 3x_2 \tag{3.5}$$

при следующих ограничениях

$$\begin{gathered} 2x_1 + 3x_2 \le 18\,; \\ x_1 - 4x_2 \le 4\,; \\ -2x_1 + x_2 \le 2\,; \\ x_1 \ge 0\,;\ x_2 \ge 0 \end{gathered} \tag{3.6}$$

Область допустимых значений (область определения) $OABCD$ для модели (3.5) образована ограничениями (3.6) (рис. 3.2). Поверхность отклика представляет собой плоский многоугольник $OA'B'C'D'$ (рис. 3.2, *б*).

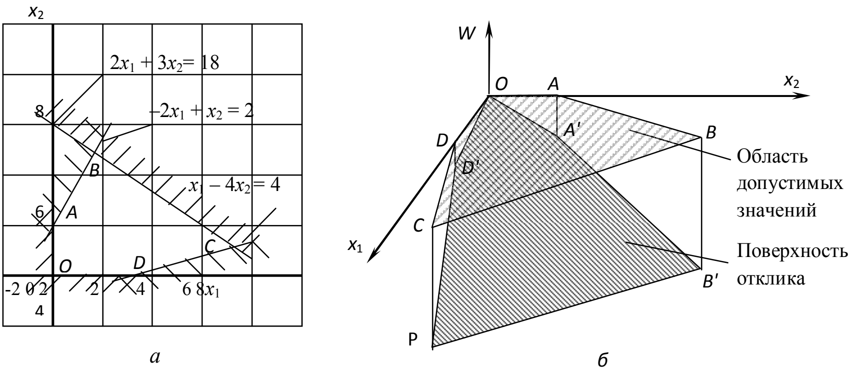


Рис. 3.2. Возможные области допустимых значений поверхностей отклика

При определенном соотношении ограничений множество допустимых решений может отсутствовать (пусто).

Пример такого множества показан на рис. 3.3.

Прямые AC и DC ограничивают область допустимых значений сверху. Третье ограничение отсекает область допустимых значений снизу от прямой AB.

Таким образом, общей области, удовлетворяющей всем трем ограничениям, нет.

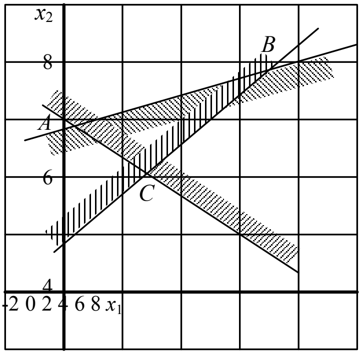


Рис. 3.3. Множество допустимых решений

Линейные модели достаточно просты и поэтому, с одной стороны, предполагают существенное упрощение задачи, а с другой – допускают разработку простых и эффективных методов решения.

Линейные модели могут использоваться при поэтапной аппроксимации нелинейных моделей (линеаризация задачи). Особенно эффективен этот прием при изучении небольших областей исследуемого пространства. Представление отдельных участков нелинейной поверхности отклика линейной моделью лежит в основе большой группы методов оптимизации, так называемых методов с линейной тактикой.

Исследование линейных моделей не представляет труда. В частности, влияние каждой из переменных на характеристики модели вида $W = a_0 + a_1 x_1 + ... + a_k x_k$, задается ее коэффициентами:

$$\frac{\partial W}{\partial x_i} = a_i, \ i = 1, 2, ..., k .$$

Для нахождения оптимума линейной модели $W_{опт}$ разработан эффективный симплекс-метод.

К линейным иногда сводятся простейшие модели стоимости, рассматриваемые как совокупность производимых затрат.

Примером такой модели является классическая модель стоимости перевозок (транспортная задача) (рис. 3.4).

Имеется k пунктов производства ($i = 1, 2, ..., k$) и m пунктов потребления ($j = 1, 2, ...m$) некоторого продукта. Количество продукта, произведенного в каждом из k пунктов производства, равно a_i; количество продукта, необходимого в каждом из m пунктов потребления, равно b_j.

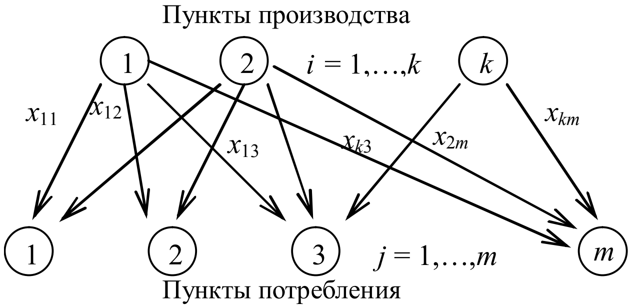


Рис. 3.4. Модель решения транспортной задачи

Предполагается равенство общего производства и потребления:

$$\sum_{i=1}^{k} a_i = \sum_{j=1}^{m} b_j \ .$$

Количество продукта, перевозимого из i-го пункта производства в j-й пункт потребления, равно x_{ij}; стоимость перевозки единицы этого продукта – c_{ij} .

Суммарная стоимость перевозок C_Σ задается линейной моделью:

$$C_\Sigma = \sum_{i=1}^{k} \sum_{j=1}^{m} c_{ij} x_{ij}$$

при следующих ограничениях

$$\sum_{j=1}^{m} x_{ij} = a_i; \quad \sum_{i=1}^{k} x_{ij} = b_j; \quad x_{ij} \geq 0.$$

К линейным также относятся модели в виде линейных дифференциальных уравнений (обыкновенных или в частных производных).

Линейное обыкновенное дифференциальное уравнение n-го порядка имеет вид:

$$a_n \frac{d^n x(t)}{dt^n} + a_{n-1} \frac{d^{n-1} x(t)}{dt^{n-1}} + \ldots + a_1 \frac{dx(t)}{dt} + a_0 x(t) = f(t) \ . \qquad (3.7)$$

Начальные условия записываются таким образом:

$$x(0) = C_0, \quad x'(0) = C_1, \quad x''(0) = C_2, \ldots, \quad x^{(n-1)}(0) = C_{n-1} \, .$$

Линейное дифференциальное уравнение в частных производных имеет вид:

$$a_0 \frac{\partial \Phi(t)}{\partial t} + a_1 \frac{\partial \Phi(t)}{\partial x_1} + a_2 \frac{\partial \Phi(t)}{\partial x_2} + \ldots + a_k \frac{\partial \Phi(t)}{\partial x_k} = f(x_1, x_2, \ldots, x_k, t) .$$

Модель, заданная в виде дифференциального уравнения в частных производных, включает начальные и граничные условия (условия на границе области определения функции $\Phi(t)$).

3.3.5. Выбор метода моделирования

Математическая модель может быть исследована различными методами – аналитическими или имитационными. В некоторых случаях наличие имитационной модели делает возможным применение математических методов оптимизации. Для использования аналитических методов необходимо математическую модель преобразовать к виду явных аналитических зависимостей между характеристиками и параметрами системы и внешних воздействий.

При имитационном моделировании динамические процессы системы – оригинала подменяются процессами, имитируемыми в абстрактной модели, но с соблюдением таких же соотношением длительностей и временных последовательностей отдельных операций. Поэтому метод имитационного моделирования мог бы называться алгоритмическим или операционным.

Методы имитационного моделирования различаются в зависимости от класса исследуемых систем, способа продвижения модельного времени и вида количественных переменных параметров системы и внешних воздействий. Методы моделирования разрабатываются для дискретных, непрерывных и смешанных дискретно – непрерывных систем.

В зависимости от способа продвижения модельного времени методы моделирования подразделяются на методы с приращением временного интервала и методы с продвижением времени до особых состояний.

В первом случае модельное время продвигается на некоторую величину Δt. Далее определяются изменения состояний элементов и выходных воздействий системы, которые произошли за это время. После этого модельное время снова продвигается на величину Δt, и процедура повторяется. Так продолжается до конца периода моделирования T_m. Этот метод моделирования называют «принципом Δt».

Во втором случае в текущий момент модельного времени t сначала анализируются те будущие особые состояния – поступление дискретного входного воздействия, завершение обслуживания и т.п., для которых определены моменты их наступления $t_i > t$. Выбирается наиболее раннее особое состояние, и модельное время продвигается до момента наступления этого состояния. Затем анализируется реакция системы на выбранное особое состояние.

В частности, в ходе анализа определяется момент наступления нового особого случая. Затем анализируются будущие особые состояния, и модельное время продвигается до ближайшего. Так продолжается до завершения периода моделирования T_m. Данный метод называют «принципом особых состояний». Метод может использоваться, только когда имеется возможность определения моментов наступления будущих очередных особых состояний.

Параметры системы и внешних воздействий могут быть детерминированными или случайными. По этому признаку различают детерминированное и статистическое моделирование. При статистическом моделировании для получения достоверных вероятностных характеристик процессов функционирования системы требуется их многократное воспроизведение с различными конкретными значениями случайных факторов и статистической обработкой результатов измерений. В основу статистического моделирования положен метод статистических испытаний, или метод Монте–Карло.

При разработке математической модели различают также принятую систему формализации – алгоритмический (программный) или структурный (агрегатный) подход. В первом случае процессы управляют компонентами (ресурсами) системы, а во втором – компоненты управляют процессами, определяют порядок функционирования системы.

3.3.5.1. Аналитические модели и методы

Аналитические методы исследования ВС сводятся к построению математических моделей, описывающих физические свойства элементов системы математическими объектами и отношениями между ними. При использовании аналитических методов оператор F, устанавливающий зависимость $Y = F(X)$ между характеристиками и параметрами системы, представляется совокупностью математических выражений. В таких моделях, называемых аналитическими, зависимость между характеристиками и параметрами может быть представлена в явной аналитической форме в виде выражения $y_m = f_m(x_1,...,x_N)$, решенных относительно искомых величин, или в неявной форме в виде уравнений $\phi(Y, X) = 0$, связывающих характеристики и параметры.

Как правило, свойства элементов и систем удается представить в аналитической форме, если принимаются определенные допущения о свойствах и поведении описываемых объектов: независимость одних факторов от других, линейность некоторых зависимостей, мгновенность переходов между состояниями и т.д. Если допущения соответствуют реальности, модель хорошо воспроизводит зависимость между характеристиками и параметрами. Однако во многих случаях допущения приводят к существенным отличиям модели от реального объекта, вследствие чего моделируемая зависимость существенно отличается от реальной, и характеристики представляются на модели с большой погрешностью. Так, предположение о том, что процессы обладают марковским свойством, может оказаться ошибочным, что приводит к большим погрешностям марковских моделей и даже к неверным оценкам. Основные аналитические методы теории массового обслуживания базируются на предположении, что интервалы времени между заявками входящих потоков и длительности обслуживания распределены по экспоненциальному закону. Когда это предположение выполняется, аналитические методы позволяют точно оценивать характеристики системы. Если же потоки и длительности существенно отличаются от предполагаемых, моделируемые характеристики могут сколь угодно отличаться от реальных.

Аналитические методы и модели ценны по следующим причинам.

1. Зависимости, полученные аналитическими методами, являются строго доказанными и их достоверность не вызывает сомнений, конечно с учетом принятых при выводе допущений. Поэтому аналитические зависимости используются в качестве своеобразных эталонов, с которыми сопоставляются результаты, получаемые другими методами.

2. Аналитические модели имеют большую познавательную ценность. Аналитические зависимости определяют характеристики для всей области значений параметров и несут в себе информацию о поведении соответствующих систем при любых сочетаниях параметров. На основе аналитических моделей легко определяются экстремальные и предельные значения характеристик и оцениваются эффекты от изменения параметров.

3. Аналитические модели характеризуются малыми объемами вычислений. Это свойство особенно важно при решении задач синтеза, поскольку оптимизация связана с многократными вычислениями характеристик при различных значениях параметров.

3.3.5.2. Потоки заявок

Простейший поток. При аналитическом моделировании характеристики системы вычисляются наиболее просто для потока заявок, называемого простейшим. Простейший поток – это поток заявок, который обладает следующими свойствами: 1) стационарность; 2) отсутствие последействия; 3) ординарность.

Стационарность означает постоянство вероятности того, что в течение определенного временного интервала поступит одинаковое количество заявок вне зависимости от расположения интервала на оси времени.

Отсутствие последействия заключается в том, что поступившие заявки не оказывают влияния на будущий поток заявок, т.е. заявки поступают в систему независимо друг от друга.

Ординарность означает, что в каждый момент времени в систему поступает не более одной заявки.

У простейшего потока интервалы времени τ между двумя последовательными заявками – независимые случайные величины с функцией распределения:

$$F(\tau) = 1 - e^{-\lambda\tau}. \tag{3.8}$$

Такое распределение называется экспоненциальным (показательным) и имеет плотность:

$$f(\tau) = \lambda e^{-\lambda\tau}, \tag{3.9}$$

математическое ожидание длины интервала:

$$M[\tau] = \int_0^\infty \tau f(\tau) d\tau = 1/\lambda, \tag{3.10}$$

дисперсию:

$$D[\tau] = \int_0^\infty (\tau - M[\tau])^2 f(\tau) d\tau = 1/\lambda^2 \tag{3.11}$$

и среднеквадратическое отклонение, равное математическому ожиданию. Экспоненциальное распределение характеризуется одним количественным параметром – интенсивностью.

Простейшие потоки заявок обладают следующими особенностями:

1. Сумма M независимых, ординарных, стационарных потоков с интенсивностями $\lambda_i (i - 1, ..., M)$ сходится к простейшему потоку с интенсивностью:

$$\lambda = \sum_{i=1}^{M} \lambda_i \tag{3.12}$$

при условии, что складываемые потоки оказывают примерно одинаковое малое влияние на суммарный поток.

2. Поток заявок, полученный в результате случайного разрежения исходного стационарного ординарного потока, имеющего интенсивность λ, когда каждая заявка исключается из потока с определенной вероятностью p независимо от того, исключены другие заявки или нет, образует простейший поток с интенсивностью $p\lambda$.

3. Интервал времени между произвольным моментом времени и моментом поступления очередной заявки имеет экспоненциальное распределение с таким же математическим ожиданием $1/\lambda$, что и интервал времени между двумя последовательными заявками.

4. Простейший поток создает тяжелый режим функционирования системы, поскольку, во-первых, большое число (63%) промежутков времени между заявками имеет длину меньшую, чем ее математическое ожидание $1/\lambda$, и, во-вторых, коэффициент вариации, равный отношению среднеквадратического отклонения к математическому ожиданию:

$$v = \frac{\sigma[\tau]}{M(\tau)}$$

и характеризующий степень нерегулярности потока, равен единице, в то время как у детерминированного потока коэффициент вариации $v = 0$, а для большинства законов распределения $0 < v < 1$.

Простейший поток имеет широкое распространение не только из-за аналитической простоты, связанной с ним теории, но и потому, что большое количество реально наблюдаемых потоков статистически не отличимы от простейшего. Этот эмпирический факт подтвержден рядом математических моделей, в которых при довольно общих условиях доказывается, что поток близок к простейшему.

Пуассоновский поток. Пуассоновским потоком называется ординарный поток заявок с отсутствием последействия, у которого число заявок, поступивших в систему за промежуток времени τ, распределено по закону Пуассона:

$$P(k,\tau) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}, \ \lambda > 0, \qquad (3.13)$$

где $P(k,\tau)$ − вероятность того, что за время τ в систему поступит точно k заявок; λ - интенсивность потока заявок.

Математическое ожидание и дисперсия распределения Пуассона равны $\lambda\tau$.

Распределение Пуассона дискретно. Стационарный пуассоновский поток является простейшим. Если нестационарный поток, интенсивность которого представляет собой функцию времени $\lambda = \lambda(t)$, описывается законом распределения Пуассона, то такой поток называется пуассоновским, но не простейшим. В распределении Пуассона длительности интервалов между двумя последовательными заявками – это случайные величины с экспоненциальным распределением.

3.3.5.3. Марковские модели

Общие сведения. В теории массового обслуживания к наиболее изученным и исследованным относятся модели, у которых случайный процесс функционирования относится к классу Марковских процессов, т.е. Марковские модели.

Случайный процесс, протекающий в системе, называется Марковским, если для любого момента времени вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент и не зависят от того, когда и как система пришла в это состояние.

При аналитическом моделировании наибольшее значение имеют Марковские случайные процессы с дискретными состояниями и непрерывным временем. Процесс называется процессом с дискретными состояниями, если его возможные состояния $s_1, s_2, \ldots$ можно заранее перечислить, т.е. состояния системы принадлежат конечному множеству, и переход системы из одного состояния в другое происходит мгновенно. Процесс называется процессом с непрерывным временем, если смена состояний может произойти в любой случайный момент.

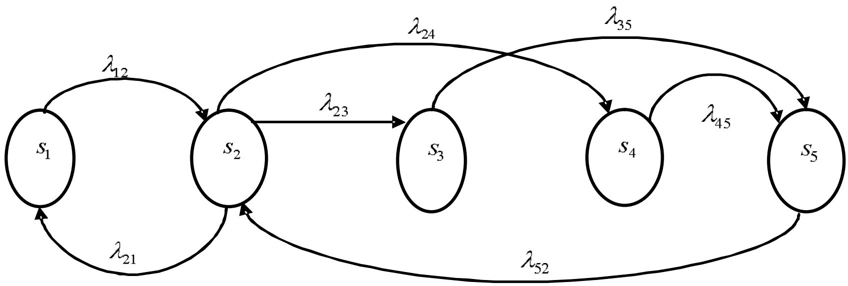


Рис. 3.5. Пример графа состояний Марковского процесса

3.3.5.4. Статистические модели

Когда отношения в объекте трудно установить из-за их многообразия, сложности и невыясненной природы процессов, используются статистические методы для математического выражения зависимостей между характеристиками и параметрами объекта. На основе эмпирических представлений о свойствах исследуемого объекта и в соответствии с целью исследования определяется состав признаков, характеризующих объект, и тип статистической модели (математические выражения, структуры). Признаками, посредством которых описывается объект, являются величины, соответствующие параметрам $x_1,...,x_N$ и характеристикам $y_1,...,y_M$ объекта. Наблюдением (измерения, регистрация) собираются статистические данные, образующие выборку $X^{(i)}=\left\{x_1^{(i)},...,x_N^{(i)}\right\}$, $Y^{(i)}=\left\{y_1^{(i)},...,y_M^{(i)}\right\}$, где $x_1^{(i)},...,x_N^{(i)}$, $y_1^{(i)},...,y_M^{(i)}$ – значения признаков при i-м наблюдении, $i=1,...,n$.

Математическая статистика предлагает обширный набор моделей и методов установления статистических закономерностей, присущих исследуемым объектам.

Наиболее широкое применение получил регрессионный анализ.

Регрессионный анализ состоит в построении функций $y_i=f_i(x_1,...,x_N)$, связывающих характеристики (зависимые переменные) с параметрами (независимыми переменными), на основе статистической выборки, содержащей статистически независимые данные. Статистическая независимость данных состоит в том, что значения признаков разных наблюдений статистической выборки не должны зависеть друг от друга. Чтобы проявились статистические зависимости, число наблюдений должно превосходить число признаков в 6 – 8 раз. Выборка должна быть однородной, т.е. относиться к объектам одного класса.

Зависимость характеристики от параметров $x_1,...,x_N$ представляется в виде линейного полинома:

$$y=b_0+b_1x_1+...+b_Nx_N,$$

а при необходимости – в виде полинома более высокого порядка

$$\begin{gathered}y=b_0+b_1x_1+...+b_Nx_N+b_{12}x_1x_2+b_{13}x_1x_3+...+\\+b_{1N}x_1x_N+...+b_{123}x_1x_2x_3+...+b_{11}x_1^2+...\end{gathered}.$$

Параметры b называются коэффициентами регрессии. Если число признаков $N=1$, то имеет место уравнение парной регрессии, при $N\ge 2$ – уравнение множественной регрессии. Переменная y рассматривается как случайная величина, которая распределена в окрестности среднего значения $\overline{y}$, зависящего от x_j, т.е.считается, что переменные x_j влияют лишь на среднее значение $\overline{y}$. Коэффициенты регрессии оцениваются по методу наименьших квадратов минимизацией дисперсии отклонения уравнения регрессии от наблюдаемых значений $y^{(i)}$, $i = 1, 2, ..., n$.

При построении регрессионной модели основными являются 2 момента:

– выбор числа независимых признаков $x_1, ..., x_N$;

– выбор формы полинома, посредством которого представляется зависимость $y = f(x_1, ..., x_N)$.

Процедуры оценки качества и улучшения моделей реализованы в пакетах прикладных программ статистического анализа регрессионных моделей. Особенностями регрессионных моделей являются следующие:

– применимость для прогноза значений y только при аргументах $x_1, ..., x_N$, принадлежащих области определения переменных, для которой построено уравнение регрессии;

– уравнения регрессии принципиально необратимы, то есть недопустимо путем тождественных преобразований из уравнения $y = f(x_1, ..., x_N)$ строить уравнение $x_i = \phi(x_1, ..., x_N, y)$, т.к. это две совершенно различные регрессии, каждая из которых должна строиться самостоятельно;

– регрессионные модели не раскрывают механизм взаимосвязи характеристик и параметров и фиксируют лишь количественную взаимосвязь величин.

Регрессионные и другие статистические модели наиболее широко используются для описания рабочей нагрузки, создаваемой прикладными задачами, а также системными процессами. Это объясняется чрезвычайной сложностью объекта. В нем совмещены свойства прикладных задач, технология обработки данных, организация операционной системы и даже конфигурация ЭВМ, для которой разрабатывается программное обеспечение. Поэтому рабочую нагрузку приходится рассматривать как черный ящик и описывать количественные взаимосвязи статистическими методами.

Регрессионные модели применяются также для компактного представления и анализа зависимостей, воспроизводимых на имитационных моделях.

3.3.5.5. Имитационные модели и методы

Методы имитационного моделирования основаны на представлении порядка функционирования системы в виде алгоритма, который называется имитационной (алгоритмической) моделью. Программа, реализующая модель, содержит процедуры, регистрирующие состояния имитационной модели и обрабатывающие зарегистрированные данные для оценки требуемых характеристик процессов и моделируемой системы.

3.3.5.6. Стохастические модели

Точные величины и зависимости, используемые в детерминированных моделях, представляют собой лишь некоторые средние значения (математические ожидания) реальных случайных величин (зависимостей). Поэтому результирующие функции, характеризующие процесс, также носят случайный характер. Результаты, полученные с помощью детерминированной модели, представляют собой математические ожидания этих характеристик. При этом конкретные данные для конкретной системы могут существенно отличаться от этих математических ожиданий. Например, ресурс конкретного двигателя может существенно отличаться от среднего ресурса двигателей данного типа. Для учета таких отличий вводятся всевозможные «запасы прочности», призванные гарантировать работоспособность реальных объектов при неблагоприятном стечении обстоятельств.

Значительно более полные и объективные результаты можно получить при переходе от детерминированных к стохастическим моделям, т. е. при переходе от точно заданных величин к соответствующим случайным величинам.

При этом константы ($\sigma, \lambda, \rho, l, ...$) заменяются случайными величинами $\xi_\sigma, \xi_\lambda, \xi_\rho, \xi_l, ...$, подчиненными определенным законам распределения.

Однократное исследование стохастической модели приведет к некоторой случайной величине функции отклика ξ_W, представляющей собой, вообще говоря, ограниченную ценность. Для получения значимых результатов необходимо провести многократное исследование модели и получить распределение результирующей характеристики в интересующем исследователя диапазоне.

Поверхность отклика в этом случае представляет собой некий размытый слой переменной плотности. Такой метод исследования стохастической модели получил название метода статистических испытаний или метода Монте-Карло.

Трудоемкость исследования стохастических моделей существенно выше, чем моделей детерминированных ввиду следующего:

1. Значительно возрастает объем исходной информации: замена констант случайными величинами, введение законов распределения этих величин усложняют модель.
2. Для получения распределения результирующей функции необходимо многократное исследование модели.

С другой стороны, полученное при статистическом моделировании распределение характеристик системы позволяет оценить не только среднее значение изучаемой величины, но и разброс этих значений, вероятности появления тех или иных значений при конкретном испытании и их зависимость от различных факторов.

Очень часто используют нормальный или гауссовский закон распределения, для которого плотность вероятности $f(x)$ и функция распределения $P(x)$ задаются следующими соотношениями:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \qquad P(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Для случайной величины ξ, распределенной по нормальному закону, $\mu = M(\xi)$, $\sigma = \sigma(\xi)$. Случайная величина распределена в интервале $\mu \pm 3\sigma$.

Наряду с нормальным используются и другие законы распределения случайных величин. Например, равномерное распределение – задает равновероятностные на отрезке $[a,b]$ случайные величины. Плотность вероятности и функция распределения при равномерном распределении определяются по формулам:

$$f(x) = \begin{cases} 0, & -\infty < x < a; \\ \dfrac{1}{b-a}, & a \le x \le b; \\ 0, & b < x < \infty. \end{cases} \qquad P(x) = \begin{cases} 0, & -\infty < x < a; \\ \dfrac{x-a}{b-a}, & a \le x \le b; \\ 1, & b < x < \infty. \end{cases}$$

Выбор закона распределения для конкретной случайной величины, входящей в стохастическую модель, может быть обоснован экспериментально или теоретически.

Конкретные параметры распределения $(\mu, \sigma, ...)$ всегда определяются на основе экспериментальных данных. При использовании метода статистических испытаний характеристики изучаемой системы оцениваются на основе некоторой ограниченной выборки реализаций. Поэтому важно определить достоверность этой оценки.

Вероятность пребывания системы в некотором состоянии (p) определяется частотой этого события при моделировании:

$$p \approx \frac{m}{n},$$

где m – число реализаций, при которых наблюдалось изучаемое состояние системы; n – общее число реализаций.

Эта оценка является приближенной, так как определяется на основе ограниченной выборки. Отношение $\frac{m}{n}$ при стохастическом моделировании рассматривается как ошибка моделирования, которая определяется отклонением выборочной статистики от вероятности $\delta = \left|\frac{m}{n} - p\right|$.

Можно показать, что эта ошибка удовлетворяет неравенству:

$$\delta \le \sqrt{\frac{p(1-p)}{\alpha n}}, \qquad (3.14)$$

Здесь p – вероятность рассматриваемого состояния; α – вероятность невыполнения оценки (3.14) (уровень риска). Доверительная вероятность выполнения этой оценки равна $1-\alpha$.

Из (3.14) следует, что погрешность стохастического моделирования обратно пропорциональна $\sqrt{n}$. То есть увеличение точности при стохастическом моделировании требует значительного увеличения числа реализаций. Для уменьшения погрешности в 10 раз необходимо увеличить число реализаций (а значит и время счета) в 100 раз. Поэтому метод статистических испытаний не может дать решения с очень высокой степенью точности. Считается, что допустимая ошибка может составлять 1-5 % максимальной величины, полученной при моделировании.

Величина ошибки зависит также от вероятности p оцениваемого состояния и допустимого уровня риска α. Обычно α задают на одном из фиксированных уровней $(\alpha = 0{,}005;\ 0{,}01;\ 0{,}025;\ 0{,}05;\ 0{,}1;\ ...)$.

3.3.5.7. Эмпирические математические модели

Переход к эмпирическим моделям предполагает заведомый отказ от аналитических методов исследования. Поэтому эмпирические модели более разнообразны и включают в себя различные по форме математические зависимости.

При разработке эмпирической математической модели предполагается использование экспериментальных данных, полученных при испытаниях объектов. Результаты таких испытаний всегда представляют собой наборы величин, характеризующих работу объекта или системы при различных сочетаниях управляющих параметров.

Наиболее эффективным средством представления результатов экспериментов в системах математического моделирования являются эмпирические модели.

При построении эмпирической модели обычно предполагается, что физическая теория работы объекта отсутствует или по тем или иным причинам не может быть использована.

Объект идентификации представляет собой так называемый «черный ящик» с некоторым числом регулируемых (или, по крайней мере, измеряемых) входов x и одним или несколькими наблюдаемыми (измеряемыми) выходами (рис. 3.6). Здесь x_i – управляющие переменные; ω_i – неопределенности (шумы); q_i – ограничения; W – характеристическая функция.

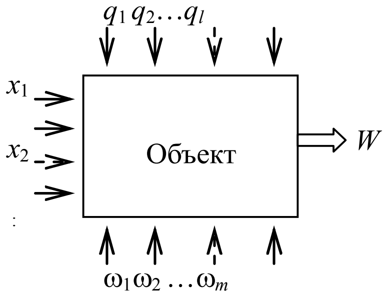


Рис.3.6. Объект идентификации

Задачей идентификации является построение модели объекта по результатам наблюдений его реакции на возмущения внешней среды.

При этом необходимо учитывать ошибки, возникающие при измерении характеристик объекта.

Требуется построить зависимость (модель)

$$W = f(x),$$

которая описывает характеристики изучаемой системы.

Это уравнение называется уравнением регрессии и описывает поверхность (гиперповерхность) отклика, характеризующую эмпирическую модель.

Обычно предполагается, что имеющиеся экспериментальные данные дают достаточно информации для воссоздания математического описания объекта.

Рассмотрим решение задачи идентификации для некоторого набора данных, полученное с помощью линейной регрессионной зависимости: $W = a + bx$. Идентификацию модели начинают с выбора формы модели, т.е. вида функции $f(x)$. При этом на практике встречаются два случая:

1. Форма математической модели известна заранее, а задача идентификации сводится к определению коэффициентов этой модели. Так, описание ряда затухающих или развивающихся процессов дается зависимостями экспоненциального типа (рис. 3.7). Задачей исследования является определение коэффициентов α, β.

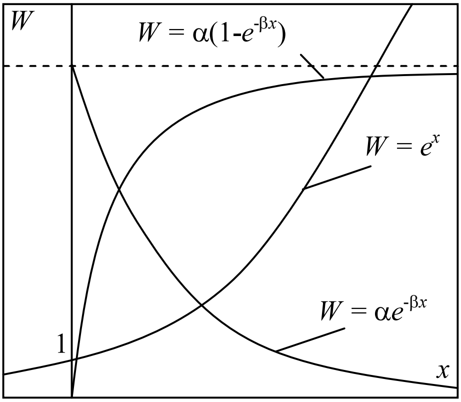


Рис. 3.7. Зависимости экспоненциального типа

2. Форма математической модели заранее неизвестна. В этом случае для идентификации модели используются отрезки бесконечных рядов, а задача заключается в определение числа членов ряда и коэффициентов при этих членах. Модель может быть представлена в виде:

$$W = \sum_{i=1}^{k} \beta_{0i} f_0(x_i) + \sum_{i=1}^{k} \beta_{1i} f_1(x_i) + \ldots + \sum_{i=1}^{k} \beta_{li} f_l(x_i),$$

где $f_q(x_i)$ – некоторые заданные функции; β_{qi} – коэффициенты регрессии; $q = 0,1,\ldots,l$.

В одномерном случае $(k = 1)$ уравнение принимает вид:

$$W = \beta_0 f_0(x) + \beta_1 f_1(x) + \ldots + \beta_l f(x).$$

Конкретный вид модели зависит от выбора функций $f_q(x)$, по которым производится разложение W. Например, при описании колебательных процессов удобно использовать ряд Фурье:

$$W = \alpha_0 + \sum_{k=1}^{¥} (\alpha_k coskx + \beta_k sinkx).$$

Часто в качестве функций $f_0(x)$, $f_1(x)$, $f_2(x),\ldots,$ $f_l(x)$ выступают степенные функции $x^0, x^1, x^2,\ldots,x^l$. Если ограничиться первыми членами разложения, то уравнения сведутся к линейным, квадратичным и другим полиномиальным моделям. Однако пока остается не ясным, сколько членов ряда обеспечивает наилучшее описание изучаемого процесса.

Обычно берут количество экспериментальных точек значительно больше, чем количество коэффициентов регрессии. В этом случае нельзя построить поверхность отклика, проходящую через все экспериментальные точки. Да этого и не требуется.

Однако можно построить приближенную модель, обеспечивающую в некотором смысле наилучшее совпадение с экспериментальными данными.

Например, прямая a построена по 10 экспериментальным точкам методом наименьших квадратов (рис. 3.8); кривая b – квадратичная модель; c – полиномиальная модель 3-го порядка достаточно хорошо соответствует исходному экспериментальному материалу, хотя проходит не через все экспериментальные точки.

Таким образом, для любой экспериментальной выборки могут быть предложены различные модели идентификации. Конкретная форма модели зависит от выбора функций $f_q(x)$ и количества членов ряда.

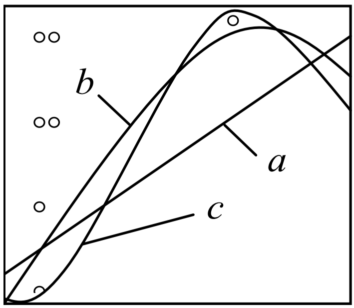


Рис. 3.8. Приближенные модели

Сама постановка задачи идентификации включает в себя элемент неопределенности, возможность множественности решений. Важно выбрать лучшее или, по крайней мере, достаточно хорошее из этих решений.

Для оценки точности модели естественно использовать величины отклонений, полученных в эксперименте величин W_j и их оценок W_{m_j}, предсказанных моделью

$$\varepsilon_j = W_j - W_{m_j} . \tag{3.15}$$

Исключительное распространение получил метод наименьших квадратов отклонений реальных значений оцениваемой величины от значений, предсказанных моделью.

Специальные методы планирования эксперимента позволяют существенно повысить объем получаемой информации, улучшают характеристики эмпирических моделей, а также упрощают процедуру обработки экспериментальных данных. Однако на практике очень часто приходится иметь дело с неорганизованным (пассивным) экспериментом. Связано это, по крайней мере, с тремя причинами:

1. Исследователь может только наблюдать входы системы, но не может их регулировать, что полностью исключает возможность планирования эксперимента.

2. Неизвестны диапазоны возможного изменения переменных (входов), что затрудняет планирование эксперимента и исключает возможность использования ряда эффективных методов планирования.

3. Приходится строить модели идентификации на основе уже полученных ранее беспорядочных экспериментальных данных.

Пример 3.2 (использование метода наименьших квадратов). Рассмотрим задачу восстановления математического описания некоторого процесса по результатам эксперимента. Предполагается, что процесс описывается одномерным уравнением 2-го порядка:

$$W = a_0 + a_1 x + a_2 x^2 \,, 0 \le x \le 6 \,.$$

Считаем, что величина x измеряется точно, а W – с ошибкой ε, имеющей нормальное распределение с нулевым математическим ожиданием и единичной дисперсией:

$$M(\varepsilon) = 0 \,, \; \sigma^2(\varepsilon) = 1 \,.$$

Выборка десяти случайных пар $(x, \tilde{W})$ представлена в табл. 3.1.

Таблица 3.1

Выборка десяти случайных пар $(x, \tilde{W})$

№	x	$\tilde{W}$	Wm	ε
1	4,8608	9,28	8,848	0,432
2	4,2396	9,40	8,821	0,579
3	2,7792	7,88	7,460	0,420
4	0,5988	1,86	2,039	-0,179
5	3,2136	7,77	8,056	-0,286
6	4,5156	8,73	8,874	-0,144
7	5,9340	8,33	8,118	0,212
8	1,5852	5,16	4,994	0,166
9	4,4880	7,28	8,872	-1,592
10	4,0932	9,22	8,767	0,453

Метод наименьших квадратов заключается в том, что неизвестные (искомые) коэффициенты a_0, a_1, a_2 должны минимизировать функцию, представляющую собой сумму квадратов невязок ε_j :

$$G = \sum_{j=1}^{10} \varepsilon_j = \sum_{j=1}^{10} (W_{m\,j} - \tilde{W}_j)^2 .$$

Минимум некоторой функции, как известно, находится в точке (a_0^*, a_1^*, a_2^*), где все частные производные этой функции по переменным a_0, a_1, a_2 равны нулю.

Для определения частных производных распишем функцию G через ее предполагаемый вид:

$$G = \sum_{j=1}^{10} (a_0 + a_1 x_j + a_2 x_j^2 - \tilde{W}_j)^2 .$$

Возьмем от функции G производные по a_0, a_1, a_2 :

$$\frac{\partial G}{\partial a_0} = \sum_{j=1}^{10} \left[2(a_0 + a_1 x_j + a_2 x_j^2 - \tilde{W}_j) \cdot 1 \right];$$

$$\frac{\partial G}{\partial a_1} = \sum_{j=1}^{10} \left[2(a_0 + a_1 x_j + a_2 x_j^2 - \tilde{W}_j) \cdot x_j \right];$$

$$\frac{\partial G}{\partial a_2} = \sum_{j=1}^{10} \left[2(a_0 + a_1 x_j + a_2 x_j^2 - \tilde{W}_j) \cdot x_j^2 \right].$$

Приравняв эти выражения к нулю и произведя некоторые преобразования, получим систему линейных алгебраических уравнений третьего порядка с тремя неизвестными, коэффициенты которой вычисляются по известным данным из табл. 3.1:

$$\begin{cases} a_0 \cdot 10 + a_1 \sum_{j=1}^{10} x_j + a_2 \sum_{j=1}^{10} x_j^2 = \sum_{j=1}^{10} \tilde{W}_j; \\ a_0 \cdot \sum_{j=1}^{10} x_j + a_1 \sum_{j=1}^{10} x_j^2 + a_2 \sum_{j=1}^{10} x_j^3 = \sum_{j=1}^{10} \tilde{W}_j x_j; \\ a_0 \cdot \sum_{j=1}^{10} x_j^2 + a_1 \sum_{j=1}^{10} x_j^3 + a_2 \sum_{j=1}^{10} x_j^4 = \sum_{j=1}^{10} \tilde{W}_j x_j^2. \end{cases}$$

Решая полученную систему, получим:

$$a_0 = -0{,}161\,;\; a_1 = 3{,}929\,;\; a_2 = -0{,}427\,.$$

Таким образом, математическая модель будет иметь вид:

$$W_m = -0{,}161 + 3{,}929x - 0{,}427x^2\,. \qquad (3.16)$$

Проверим адекватность модели методом Фишера. Для этого заполним четвертый и пятый столбцы табл. 3.1, подставляя в математическую модель (3.16) и затем в формулу (3.15) значения x_j из первого столбца.

Определим число степеней свободы системы по формуле:

$$f_s = n - m - 1\,,$$

где $n = 10$ – количество экспериментальных точек; $m = 3$ – количество неизвестных коэффициентов. То есть $f_s = 6$.

Выборочная дисперсия вычисляется по формуле:

$$s^2(\varepsilon) = \frac{1}{f_s}\sum_{j=1}^{10}(\tilde{W}_j - W_{mj})^2 = \frac{1}{6}\sum_{j=1}^{10}(\tilde{W}_j - W_{mj})^2 = 0{,}607\,.$$

Критерий Фишера вычисляется по формуле:

$$F = \frac{s^2(\varepsilon)}{\sigma^2(W)} = 0{,}607\,.$$

По статистическим таблицам при 5%-м уровне риска $(\alpha = 0{,}05)$ находим пороговое значение критерия Фишера

$$F_{f,\alpha} = F_{6;0,05} = 2{,}01\,.$$

Так как полученное значение F меньше критического (порогового), гипотеза об адекватности модели реальному процессу принимается.

3.3.6. Выбор средств моделирования и разработка программной модели

3.3.6.1. Технические средства моделирования

Современные ЭВМ позволяют моделировать сложные распределенные динамические системы. Фактор распределенности играет важную роль и предполагает построение многопроцессорных вычислительных систем на основе локальных вычислительных сетей. Поэтому для моделирования таких систем перспективным представляется использование распределенных многопроцессорных вычислительных систем или систем, эмулирующих их работу.

3.3.6.2. Алгоритмические языки общего назначения

Алгоритмические языки общего назначения применимы для реализации аналитических методов моделирования. При использовании алгоритмических языков для программирования имитационных алгоритмов возникают трудности. Первая из них заключается в том, что алгоритмы поведения сложных систем являются параллельными, то есть предполагают выполнение более чем одного преобразования в каждый момент времени. Программная имитация параллельных процессов при использовании языков общего назначения сводится к организации псевдопараллельного развития параллельных процессов, что достаточно сложно для программирования.

Вторая трудность состоит в том, что объемы данных в имитационных алгоритмах трудно оценить априорно. Поэтому требуется использовать динамическое распределение памяти, что не всегда обеспечивает язык программирования.

3.3.6.3. Языки моделирования

При создании программ имитационного моделирования возникают задачи, общие для широкого класса моделей:

– организация псевдопараллельного выполнения алгоритмов;

– динамическое распределение памяти;

– операции с модельным временем, отражающим астрономическое время функционирования оригинала;

– имитация случайных процессов;

– ведение массива событий;

– сбор и обработка результатов моделирования.

Решение перечисленных выше задач осуществляется полностью или частично внутренними средствами языка моделирования.

По структуре и правилам программирования языки моделирования подобны процедурно-ориентированным алгоритмическим языкам высокого уровня. Операторы языков моделирования выполняют более сложные процедуры, поэтому они имеют более высокий уровень и упрощают составление программ. Языки моделирования рассматриваются как формализованный базис для создания математических моделей.

3.3.6.4. Автоматизированные системы моделирования

Дальнейшее упрощение и ускорение процесса программирования привело к необходимости его автоматизации. К настоящему времени создан ряд систем автоматизации имитационного моделирования. Разработка таких систем сложное, но перспективное направления развитие средств моделирования. Наличие таких систем позволит сократить время разработки моделей, повысить качество их исследования, и проводить перспективные исследования.

3.3.7. Проверка адекватности и корректировка модели

3.3.7.1. Проверка адекватности

Адекватность модели нарушается по многим причинам: из-за идеализации внешних условий и режимов функционирования; исключения тех или иных параметров; пренебрежения некоторыми случайными факторами. Отсутствие точных сведений о внешних воздействиях, определенных нюансах структуры системы, принятые аппроксимации, интерполяции, предположения и гипотезы также ведут к уменьшению соответствия между моделью и объектом. Это приводит к тому, что результаты моделирования будут существенно отличаться от реальных.

Простейшей мерой адекватности может служить отклонение некоторой характеристики y_0 оригинала и y_m модели.

$$\Delta y = \left| y_0 - y_m \right| \quad \text{или} \quad \Delta y = \left| y_0 - y_m \right| / \, y_0 \, .$$

Можно считать, что модель адекватна с системой, если вероятность того, что отклонение Δy не превышает предельной величины Δ, больше допустимой вероятности P_Δ:

$$P\left(\Delta y < \Delta \right) \geq P_\Delta$$

Практическое использование данного критерия адекватности зачастую невозможно по следующим причинам:

– для проектируемых или модернизируемых систем отсутствует информация о значении характеристики y_0;

– система оценивается не по одной, а по множеству характеристик, у которых может быть разная величина отклонения;

– характеристики могут быть случайными величинами и функциями, а часто и нестационарными;

– отсутствует возможность априорного точного задания предельных отклонений Δ и допустимых вероятностей P_Δ.

Несмотря на это проверять адекватность необходимо. Иначе по неверным результатам моделирования будут приняты неправильные решения. На практике оценка адекватности обычно проводится путем экспертного анализа разумности результатов моделирования. Можно выделить следующие виды проверок:

– проверка моделей компонентов;

– проверка модели внешних воздействий (аппроксимации и гипотезы необходимо оценить математическими методами);

– проверка концептуальной модели функционирования системы;

– проверка формализованной и математической модели;

– проверка способов измерения и вычисления выходных характеристик;

– проверка программной модели.

3.3.7.2. Статистические методы проверки адекватности математических моделей

Если имеются или могут быть получены необходимые и достоверные экспериментальные данные, для проверки адекватности моделей можно использовать методы математической статистики.

Математически задача проверки адекватности модели формулируется как задача проверки предположения о том, что значение отклика модели W_m отличается от реального отклика системы W не более чем на заданную величину ε^*:

$$\left|W(x) - W_m\right| = \left|\varepsilon\right| \le \varepsilon^*, \quad x \in X^*. \tag{3.17}$$

Однако, истинное значение отклика системы никогда неизвестно. Полученный в результате эксперимента отклик $\tilde{W}$ в силу неконтролируемого дрейфа системы, разброса характеристик ее элементов и, наконец, просто ошибок измерения представляет собой случайную величину, отличающуюся от W. Поэтому при сравнении результатов математического и физического экспериментов $(W_{m\,i}, \tilde{W}_i)$ будет получена совокупность случайных величин $\{\varepsilon_i\}: \tilde{W}_i(x) - W_{m\,i}(x) = \varepsilon_i, \quad i = 1,...,n$,среди которых могут оказаться как величины, удовлетворяющие условию (3.15), так и не удовлетворяющие ему.

Для объяснения, являются ли полученные отклонения $\left(\varepsilon_i > \varepsilon^*\right)$ случайными причинами или их наличие должно быть признано существенным, что приводит к отказу от проверяемой модели, на основе выборки случайных величин $\{\varepsilon_i\}$ строят статистические критерии, по которым оценивают адекватность модели.

Гипотеза об адекватности модели действительности (гипотеза H_0) может быть сформулирована как предположение о том, что полученная совокупность $\{\varepsilon_i\}$ не дает оснований отказаться от рассматриваемой модели. Иными словами, модель удовлетворяет заданной точности ε^*.

Альтернативная гипотеза H_1 состоит в том, что модель не отвечает заданным требованиям (3.17) и, следовательно, должна быть отвергнута.

Так как выборка $\{\varepsilon_i\}$ случайна, решение о выборе одной из гипотез H_0 или H_1 носит вероятностный характер. При этом может быть допущена ошибка первого рода, состоящая в отказе от правильной модели (принимается H_1, когда верна H_0), или ошибка второго рода, состоящая в принятии ошибочной модели (принимается H_0, когда верна H_1). Вероятность ошибки первого рода обозначают через α, второго рода – β. Принято называть α риском разработчика, β – риском потребителя. Разумеется, желательно минимизировать как α, так и β. Однако, при заданном объеме экспериментальной выборки уменьшение α влечет за собой увеличение β. На практике α задается на определенном уровне $(\alpha = 0{,}05; 0{,}01; 0{,}005; 0{,}001)$.

Величина $1-\beta$ характеризует вероятность отказа от ошибочной модели, называется мощностью критерия и является мерой его эффективности. Выбор вероятностей ошибок α и β при проверке конкретной модели зависит от ответственности решений, принимаемых на основе моделирования.

Для оценки гипотезы об адекватности модели используют несколько критериев:

1) критерий согласия χ^2 Пирсона;

2) критерий Смирнова-Колмогорова;

3) критерий Фишера и др.

Критерий Смирнова-Колмогорова целесообразно использовать при относительно малых выборках, когда критерий χ^2 оказывается неэффективным.

Критерий Фишера реализован на основе анализа дисперсий. Если дисперсия, характеризующая ошибку эксперимента $\sigma^2(W)$, известна, вычисляется выборочная дисперсия $S^2(\varepsilon)$ и составляется F-отношение:

$$F_{f_S,\infty} = \frac{S^2(\varepsilon)}{\sigma^2(W)}.$$

Полученную величину F-отношения сравнивают с пороговым значением критерия Фишера $F_{f \cdot s,\infty,\alpha}$ при заданном уровне риска α.

При $F_{f \cdot s,\infty} \le F_{f \cdot s,\infty,\alpha}$ полученная величина $S^2(\varepsilon)$ может быть объяснена случайным разбросом экспериментальных данных и, следовательно, нет оснований для отказа от проверяемой модели.

Если $F_{f \cdot s,\infty} > F_{f \cdot s,\infty,\alpha}$, полученное расхождение результатов моделирования и экспериментальных данных зн?чимо и, следовательно, модель должна быть отвергнута как недостаточно точная.

3.3.7.3. Выбор оптимальной эмпирической модели

Принцип наименьших квадратов позволяет найти наилучшую модель идентификации для исследуемой экспериментальной выборки с заданным уравнением регрессии вида:

$$W = \sum_{i=1}^{k} \beta_{0i} f_0(x_i) + \sum_{i=1}^{k} \beta_{1i} f_1(x_i) + \ldots + \sum_{i=1}^{k} \beta_{li} f_l(x_i). \qquad (3.18)$$

Если имеются достаточно веские основания для выбора формы этого уравнения, никаких проблем не возникает. Однако в большинстве случаев конкретная форма модели заранее неизвестна и может, вообще говоря, быть различной.

Может показаться, что более сложная модель (увеличение степени полинома) всегда обеспечивает получение б?льшей точности. На самом деле это не так. При переходе к полиномам более высокой степени можно, получить лучшее согласие регрессионной кривой с экспериментальными данными. Для $m = n$ это согласие будет абсолютным, но при этом получится худшее согласие с истинным характером процесса $W(x)$. Дело в том, что экспериментальные данные представляют собой случайные величины и содержат лишь ограниченную информацию о характере $W(x)$. Увеличение степени полинома целесообразно лишь до тех пор, пока из экспериментальной выборки извлекается надежная информация. Таким образом, возникает проблема выбора формы модели.

Подход к решению этой проблемы основан на статистическом исследовании уравнений регрессии.

1. Метод всех возможных регрессий основан на последовательном изучении всех возможных моделей $(m < n)$, из которых отбирается лучшая модель. Метод представляется мало пригодным для анализа сложных систем, так как отличается высокой трудоемкостью.

2. Метод исключения предполагает исследование наиболее полной (в пределах разумного) модели и последовательную проверку на значимость всех ее членов. При этом для каждого из членов модели вычисляется величина критерия Фишера F. На основе полученного множества $\{F_i\}$ выбирается член уравнения регрессии, соответствующий минимальному значению критерия F_i. Если это минимальное значение меньше критического при выбранном уровне риска $\left(F_i < F_{kp,\alpha}\right)$, то соответствующий член исключается из регрессионного уравнения как несущественный, после чего все коэффициенты регрессии пересчитываются заново и вновь осуществляется проверка их значимости.

Если $\left(F_i > F_{kp,\alpha}\right)$,то все члены модели существенны и уравнение регрессии остается в первоначальном виде. Если это произошло уже на первом шаге исследования, ст?ит рассмотреть целесообразность усложнения первоначальной модели. Трудоемкость этого метода меньше, чем метода всех возможных регрессий.

3. Метод включений по существу противоположен методу исключений и предусматривает последовательное включение в модель новых членов с проверкой их статистической значимости.

Трудоемкость этого метода существенно меньше трудоемкости рассмотренных выше методов.

Общим недостатком всех рассмотренных выше методов является использование для оценки модели того же экспериментального материала, на основе которого эта модель построена.

4. Использование регуляризации, предполагает что все экспериментальные данные разбиваются на две части: обучающую$\left(n_1\right)$и проверочную $\left(n_2\right)$. Первая из них используется для определения коэффициентов регрессии модели, вторая – для оценки модели в целом.

Оптимальные по этому подходу модели мало чувствительны к небольшим изменениям исходных данных.

Число точек обучающей последовательности должно быть, по крайней мере, на единицу больше числа коэффициентов регрессии $\left(n_1 > m+1\right)$.Для повышения достоверности результатов этот запас должен быть существенно увеличен $\left(n_1 \geq (2...3)m\right)$. Проверочная последовательность должна включать в себя хотя бы одну точку.

В ряде случаев в качестве критерия регуляризации удобно использовать критерий несмещенности, обеспечивающий наименьшее изменение модели при изменении состава обучающей последовательности. При этом весь экспериментальный массив разбивается на две одинаковые по величине последовательности $(n_1 = n_2)$, каждая из которых поочередно используется в качестве обучающей. В результате их использования определяются две независимые, одинаковые по форме модели $W_m^*(n_1)$ и $W_m^{**}(n_2)$. Оптимальная модель ищется по всем точкам выборки. При этом, критерий регуляризации всегда имеет четко выраженный минимум, что обеспечивает объективное выделение модели оптимальной сложности.

3.3.7.4. Корректировка модели

Если по результатам проверки адекватности выявляется недопустимое рассогласование модели и системы, возникает необходимость в корректировке или калибровке модели. При этом выделяются следующие типы изменений: глобальные, локальные и параметрические.

Стратегия корректировки модели должна быть направлена на первоочередное введение глобальных изменений, затем – локальных и параметрических изменений. Для выработки тактики параметрических изменений большое значение имеет анализ чувствительности модели к вариациям ее параметров.

Завершается этап проверки адекватности и корректировки модели определением и фиксацией области пригодности модели. Под областью пригодности понимается множество условий, при соблюдении которых точность результатов моделирования находится в допустимых пределах.

3.3.8. Планирование экспериментов с моделью

3.3.8.1. Стратегическое планирование

Цели моделирования достигаются путем исследования разработанной модели. Исследования заключаются в проведении экспериментов по определению выходных характеристик системы при разных значениях управляемых переменных параметров модели. Эксперименты следует проводить по определенному плану.

Особо важным является планирование экспериментов при численном и статистическом моделировании. Это обусловлено большим числом возможных сочетаний значений управляемых параметров, а каждый эксперимент проводится при определенном сочетании значений параметров. Поэтому возникает необходимость в выборе определенных сочетаний параметров и последовательности проведения экспериментов. Это называется стратегическим планированием.

Разработка плана начинается на ранних этапах создания модели, когда выявляются характеристики качества и параметры, с помощью которых предполагается управлять качеством функционирования системы. Эти параметры называют факторами. Затем выделяются возможные значения количественных параметров и варианты качественных (функциональных) параметров. Их называют уровнями q. При этом число сочетаний равно

$$N_s = q_1 \times q_2 \times ... \times q_k ,$$

где k – число факторов.

Если число факторов велико, то для проведения исследований системы используется один из методов составления плана по неполному факторному анализу. Эти методы хорошо разработаны в теории планирования экспериментов.

3.3.8.2. Тактическое планирование

Совокупность методов уменьшения длительности машинного эксперимента при обеспечении статистической достоверности результатов имитационного моделирования получила название тактического планирования. На длительность одного эксперимента (периода моделирования T_m) влияет степень стационарности системы, взаимозависимости характеристик и значения начальных условий моделирования.

Данные, собранные в эксперименте, можно рассматривать как временные ряды, состоящие из замеров определенных характеристик. Ряд замеров характеристики y может рассматриваться как выборка из стохастической последовательности. Если эта последовательность стационарна, ее среднее $\overline{y}$ не зависит от времени. Оценкой $\overline{y}_N$ является среднее по временному ряду $y_1,...,y_N$. Для эргодической последовательности точность этой оценки возрастает с ростом N.

Если заданы максимальная допустимая ошибка оценки (доверительный интервал) и минимальная вероятность того, что истинное среднее $\overline{y}$ лежит внутри этого интервала, то существует минимальный размер исследуемой выборки. Этот размер соответствует минимальной длительности эксперимента. Для оценки нескольких характеристик период моделирования определяется по максимальному значению.

3.3.9. Моделирование на ЭВМ и анализ результатов моделирования

3.3.9.1. Обработка результатов измерений имитационного моделирования

При статистическом моделировании в ходе имитационного эксперимента измеряются множества значений по каждой выходной характеристике. Эти выборки необходимо обрабатывать для удобства последующего анализа и использования. Поскольку выходные характеристики зачастую являются случайными величинами или функциями, обработка заключается в вычислении оценок математических ожиданий, дисперсий и корреляционных моментов.

Для того, чтобы исключить необходимость хранения результатов всех измерений, обработку проводят по рекуррентным формулам, когда оценки вычисляют в процессе эксперимента методом нарастающего итога по мере появления новых измерений.

Для стохастических характеристик можно построить гистограмму относительных частот – эмпирическую плотность распределения. С этой целью область предполагаемых значений характеристики y разбиваются на интервалы. В ходе эксперимента по мере измерений определяют число попаданий характеристики в каждый интервал и подсчитывают общее число измерений. После завершения эксперимента для каждого интервала вычисляют отношение числа попаданий характеристики к общему числу измерений и длине интервала.

Для случайных нестационарных характеристик период моделирования T_m разбивается на отрезки Δt и запоминаются значения характеристики в конце каждого Δt. Проводится серия экспериментов с разными последовательностями случайных параметров модели. Измерения по каждому Δt обрабатываются как при оценке случайных величин.

3.3.9.2. Определение зависимостей характеристик от параметров системы

Для анализа зависимостей характеристик от параметров системы и внешних воздействий можно воспользоваться корреляционным, дисперсионным или регрессионным методами.

С помощью корреляционного анализа можно установить наличие связи между двумя или более случайными величинами. Оценкой связи служит коэффициент корреляции при наличии линейной связи между величинами и нормальном законе их совместного распределения.

Дисперсионный анализ можно использовать для установления относительного влияния различных факторов на значения выходных характеристик. При этом общая дисперсия характеристики разлагается на компоненты, соответствующие рассматриваемым факторам. По значениям отдельных компонентов делают вывод о степени влияния того или другого фактора на анализируемую характеристику.

Когда все факторы в эксперименте являются количественными, можно найти аналитическую зависимость между характеристиками и факторами. Для этого используются методы регрессионного анализа.

К анализу результатов моделирования можно отнести задачу анализа чувствительности модели к вариациям ее параметров. Под анализом чувствительности понимают проверку устойчивости характеристик процесса функционирования системы к отклонениям значений параметров.

3.3.9.3. Использование результатов моделирования

Результаты моделирования используются для принятия решения о работоспособности системы, для выбора лучшего проектного варианта или для оптимизации системы. Решение о работоспособности принимается по тому, выходят или не выходят характеристики системы за установленные границы при любых допустимых изменениях параметров. При выборе лучшего варианта из всех работоспособных вариантов выбирается тот, у которого максимальное значение критерия эффективности. Наиболее общей и сложной является оптимизация системы: требуется найти такие сочетания значений переменных параметров системы или рабочей нагрузки из множества допустимых, которое максимизирует значение критерия эффективности:

$$E_{opt} = max\left(min\right)E\left(y\right),$$

при соблюдении ограничений на все n характеристик

$$y_{i_{\min}} \le y_i \le y_{i_{\max}}, \; i = 1, 2, ..., n .$$

Если выходная характеристика y_i является случайной величиной с некоторой плотностью распределения $f\left(y_i\right)$, целесообразно ввести в задачу оптимизации стохастические ограничения следующего вида:

$$P\left(y_{i_{\min}} \le y_i \le y_{i_{\max}}\right) = \int_{y_{i_{\min}}}^{y_{i_{\max}}} f\left(y_i\right) dy_i \ge \beta_i ,$$

где β_i – минимально допустимая вероятность того, что конкретные значения y_i не выйдут за ограничивающие пределы.

3.4. Пример практической реализации и исследования модели деятельности оператора и использованием имитационного моделирования

В главе 2 представлена модель деятельности оператора с описанием содержания ее элементов (подраздел 2.7, рис. 2.13, 2.14).

В результате имитационного моделирования представленной модели были получены следующие данные. (рис. 3.9 – 3.18).

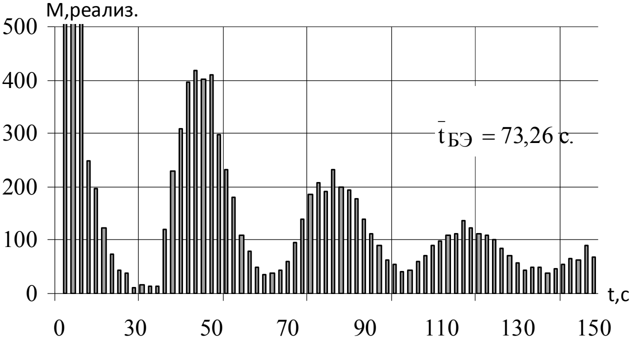


Рис. 3.9. Затраты времени оператора на оценку информации о ВО, представленной на БЭ

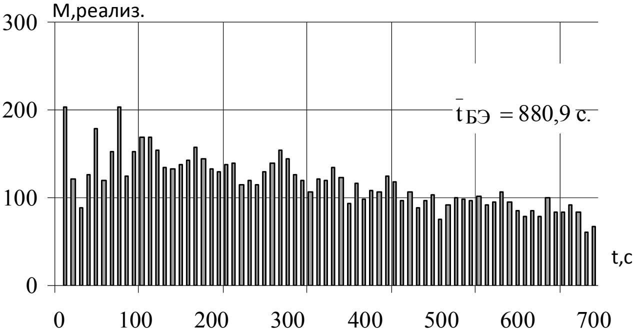


Рис. 3.10. Затраты времени оператора на оценку информации о ВО, представленной на БЭ, в случае самостоятельной оценки ВО

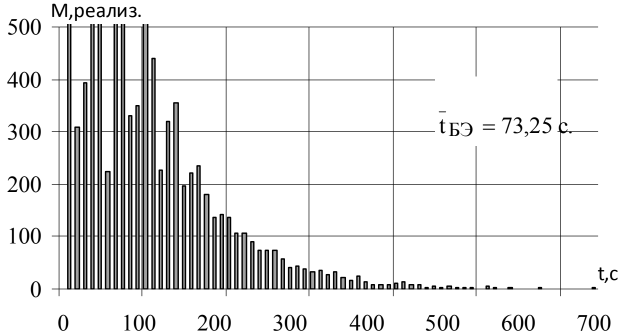


Рис. 3.11. Затраты времени оператора на оценку информации о ВО при взаимодействии с другими операторами

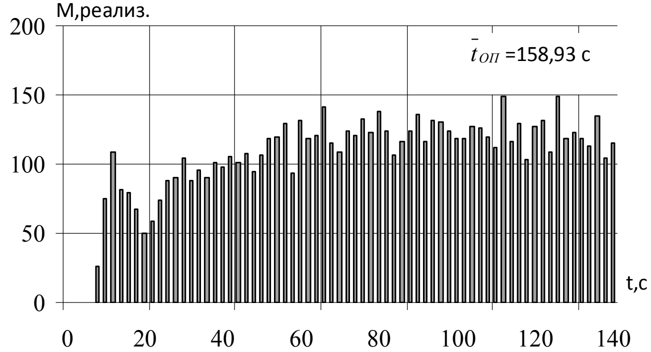


Рис. 3.12. Распределение затрат времени на работу оператора с другими операторами

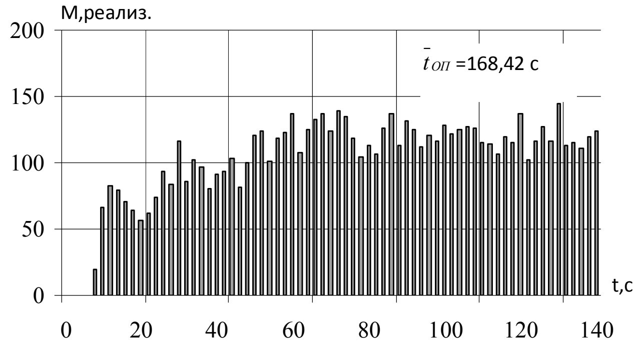


Рис. 3.13. Распределение затрат времени на работу оператора в простых условиях

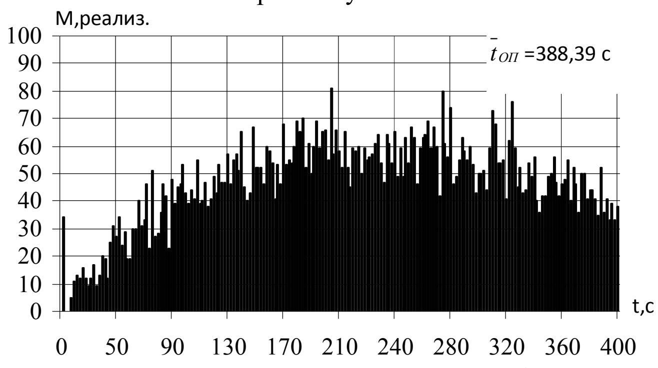


Рис. 3.14. Распределение затрат времени на работу оператора при оценке сложной обстановки

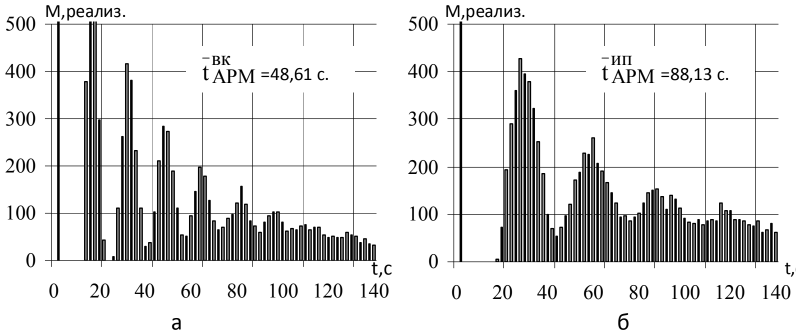


Рис. 3.15. Затраты времени: а – на ввод команд в ЭВМ;
б – на оценку информации на СОИ АРМ

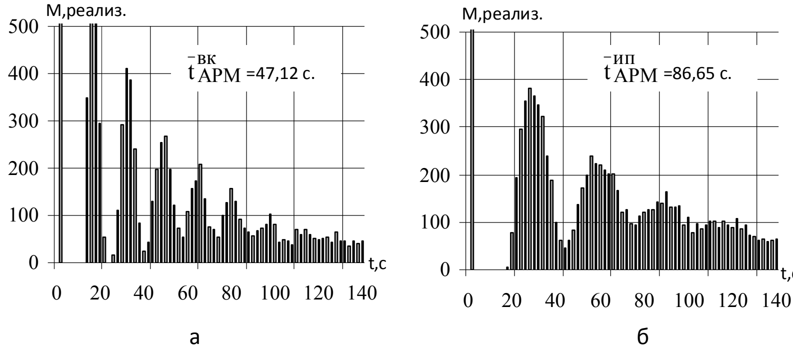


Рис. 3.16. Затраты времени: а – на ввод команд в ЭВМ;
б – на оценку информации на СОИ АРМ в простых условиях

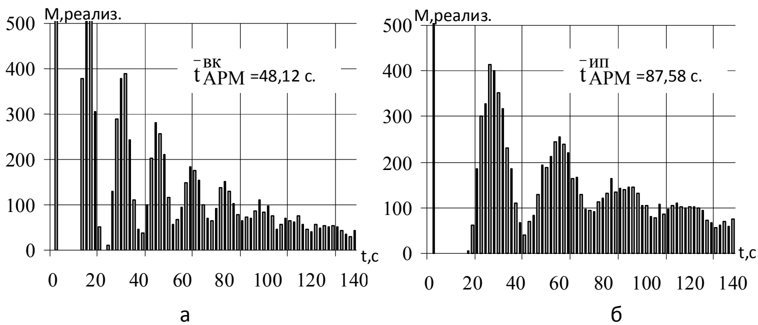


Рис. 3.17. Затраты времени: а – на ввод команд в ЭВМ;
б – на оценку информации на СОИ АРМ в сложных условиях

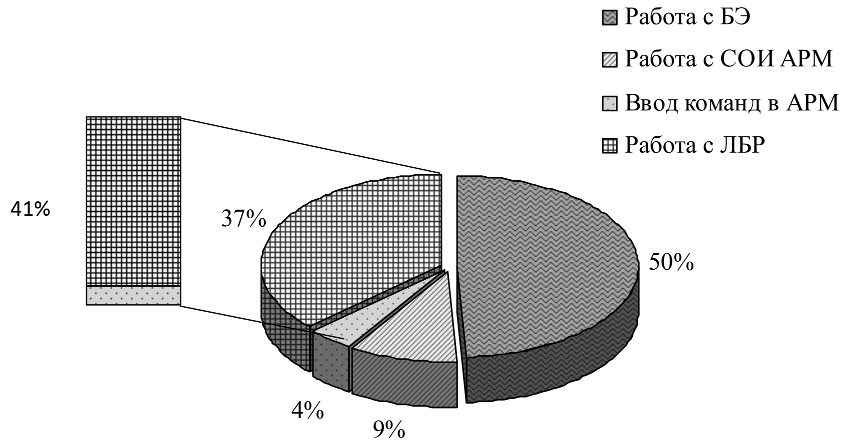


Рис. 3.18. Распределение затрат времени оператора при оценке воздушной обстановки

В данном случае проведенное имитационное моделирование позволяет получить целые ряд временных характеристик деятельности оператора по оценке воздушной обстановки как в простых, так и в сложных условиях. Анализ полученных на основе моделирования данных позволяет оценить напряженность работы оператора в различных условиях обстановки, степень сложности работы с различными техническими устройствами, выявлять наиболее сложные операции в процессе оценки воздушной обстановки.

ГЛАВА 4. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Изучение вопросов работы человеко-машинных систем очень и очень часто сталкивается с проблемой обработки информации оператором. При этом интересен даже не сам процесс обработки, а то, сколько этой самой информации обрабатывает оператор. Попытки использования классических методов оценки объема информации, как правило, к успеху не приводят. Оцененные с их помощью размеры информации оказываются зачастую просто огромными, такими, что оператор теоретически не может с ними справляться. На самом деле, в процессе работы он отлично справляется с существующими потоками информации, и это не вызывает в ходе работы особых трудностей. В чем же дело? В ходе экспериментов стало ясно, что классические методы подсчета количества информации зачастую не работают. Они не работают особенно тогда, когда речь идет о количестве обработанной оператором информации, необходимом количестве информации для принятия решений оператором, подсчете информационной емкости устройств отображения информации, подсчете информационной насыщенности информационных моделей и ряде других задач, возникающих при эргономическом проектировании систем информационного обеспечения деятельности операторов.

Данная глава представляет собой реферативный материал на тему неклассических методов подсчета количества информации, который может оказаться полезным при решении описанных выше задач. Для полноты описания здесь же в качестве справочного материала приведены и основные положения классической теории информации. Однако основной задачей данной главы видится создание почвы для поиска новых и пересмотра существующих подходов к подсчету количества информации в процессе деятельности оператора, в ходе обработки информации и принятия решений, оценки информационных возможностей систем информационного обеспечения деятельности операторов, при эргономическом проектировании систем информационного обеспечения.

4.1. Общие аспекты теории обработки информации

Теория обработки информации изучает то, как люди обращаются с информацией, отбирают и усваивают ее, а затем используют в процессе принятия решений и управления своим поведением. Психологи, изучающие процессы обработки информации, строят теории когнитивных способностей и поведения человека, используя понятия из области вычислительной техники, лингвистики и теории информации. Такой подход стимулировал разработку жизнеспособных теорий и значимых исследований в различных областях психологии, особенно касающихся изучения восприятия, памяти, внимания, речи, мышления и решения задач. Процессы обработки информации в настоящее время являются ведущим ориентиром в таких областях знаний как эргономика, экспериментальная психология и информационные технологии.

Перспективная методология изучения процессов обработки информации представляет собой семейство разрозненных теоретических и исследовательских программ. Здесь существует лишь частичное согласие по исходным посылкам, теории и методам исследования. Тем не менее, имеется достаточное число совпадающих моментов для описания этого семейства как парадигматической подгруппы, четко отличающейся от близких, таких как трансформационная лингвистика и психология Пиаже, или более отдаленных подходов, таких как радикальный бихевиоризм.

Психологию обработки информации можно разделить на элементы, возникшие внутри экспериментальной психологии, и элементы, заимствованные из внешних дисциплин. Вклады других дисциплин включают математическую логику, технику связи и теорию информации, трансформационную лингвистику и теорию вычислительных систем. Некоторые из этих предшествующих направлений и дисциплин были положительно восприняты в целом или частично.

4.1.1. Влияние бихевиоризма середины XX века

Бихевиоризм (от англ. behaviour – поведение) – психологическое направление, начало которого было положено публикацией в 1913 г. статьи американского психолога Дж. Уотсона «Психология с точки зрения бихевиориста». В качестве предмета психологии в нем фигурирует не субъективный мир человека, а объективно фиксируемые характеристики поведения, вызываемого какими-либо внешними воздействиями. При этом, в качестве единицы анализа поведения постулируется связь стимула (S) и ответной реакции (R). Все ответные реакции можно разделить на наследственные (рефлексы, физиологические реакции и элементарные «эмоции») и приобретенные (привычки, мышление, речь, сложные эмоции, социальное поведение), которые образуются при связывании (обусловливании) наследственных реакций, запускаемых безусловными стимулами, с новыми (условными) стимулами. В частности, в исследованиях Уотсона показано, что, если сочетать безусловные стимулы, вызывающие у младенца эмоцию страха (резкий звук, потеря опоры), с другими, первоначально нейтральными (например, показ белого кролика), то через некоторое время реакция страха может вызваться уже одним только показом кролика. Но в дальнейшем было показано, что само обусловливание представляет собой достаточно сложный процесс, имеющий психологическое содержание. Постепенно возникли изменения в концептуальном аппарате бихевиоризма, что заставило говорить о преобразовании его в необихевиоризм. В схеме $S-R$ появились «промежуточные переменные» (образ, цель, потребность). Другим вариантом ревизии классического бихевиоризма стала концепция оперантного бихевиоризма Б. Скиннера, разработанная в 30-х гг. XX в., где было модифицировано понятие реакции. В целом, бихевиоризм оказал большое влияние на развитие психотерапии и методы программированного обучения.

Бихевиоризм был в основе своей антименталистским, занимался изучением поведения животных и активно распространял принципы обусловливания на все области психологии. Препятствия, возникшие при попытке распространить теорию и метод бихевиоризма на символические процессы человека, в частности, на языковые способности – стали одним из основных факторов, стимулировавших возникновение парадигмы обработки информации. Когда разочарование стало всеобщим, психологи обратились к другим теориям для ориентации своих исследовательских программ. Результатом явился существенный отход от бихевиоризма.

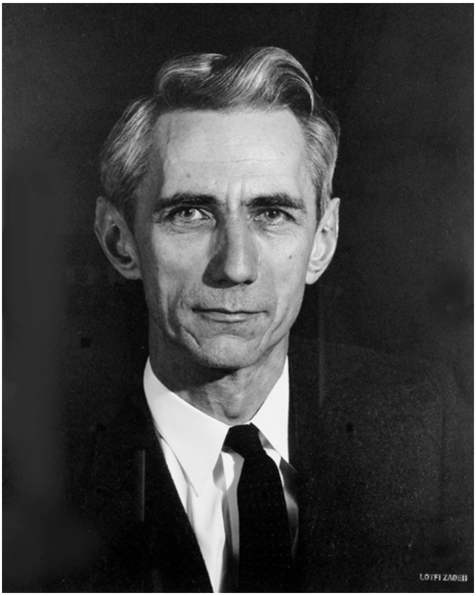
Психологи, работающие в парадигме обработки информации, все еще разделяют со своими предшественниками - бихевиористами твердую веру в эмпиризм, операционализм и номотетический идеал. Поэтому общие методологические и статистические предпочтения при смене парадигм сохранились. Психологи приняли новые способы рассмотрения привычных вопросов и модернизировали планы своих экспериментов.

Новый взгляд отверг некоторые аспекты бихевиористской традиции: всеобщую экстраполяцию ограниченного набора принципов научения; получаемые на животных данные как источник основных принципов; и обусловливание как основную форму обучения. Возможно, наиболее важно то, что психологи, придерживающиеся парадигмы обработки информации, отказались от антименталистской установки бихевиоризма, вместе с его крайним энвайронментализмом и исключительным акцентом на внешних причинах поведения. Снова были признаны врожденные способности и предложены к обсуждению внутренние процессы, такие как планы, стратегии, образы и решения. Эксперименты с людьми в качестве испытуемых пришли на смену экспериментам с животными, а внутренние процессы были увязаны с внешними детерминантами поведения. В конце концов, большинство исследователей, занимавшихся психическими процессами более высокого уровня, обнаружили, что внутренние процессы и структурные компоненты являются достаточно приемлемыми элементами теории. Эта переориентация была в значительной степени облегчена аналогиями, фактами, методами и теориями из родственных наук, которые в том или ином аспекте занимались символами и их обработкой.

4.1.2. Влияние теории передачи информации

Ошеломляющее развитие в XX веке техники связи – телефонии, радио и телевидения – произошло потому, что уже было известно многое о теоретической сущности систем связи. Ученые-связисты сформулировали и подвергли проверке общие законы, описывающие обобщенные режимы работы идеальных и реальных систем связи. Понятия и законы, пригодные для характеристики систем связи, применялись к биологическим и физическим системам, обычно не рассматриваемым в качестве информационных каналов.

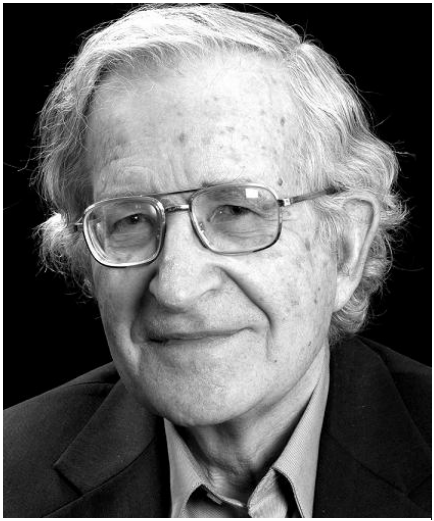
На раннем этапе развития на психологов, работавших с парадигмой обработки информации, произвела большое впечатление возможная аналогия человеческого информационного процессора с описанным в теории связи информационным каналом. Некоторые способности человека представлялись в виде части информационно-управляющего канала, обладающего внутренними состояниями и ограничениями. Были спланированы исследования для определения свойств канала и способности к передаче различных типов информации. Ряд технических понятий, таких как “неопределенность”, “информация”, “бит”, “источник”, “сообщение”, “адресат” и “кодирование”, был позаимствован из теории связи. Главным импульсом для технологического и концептуального переноса достижений теории и техники связи на сферу экспериментальной психологии оказалась работа Клода Шеннона. Созданная Шенноном математическая теория информации была применима к любому сообщению из любого источника, передавае-



Клод Элвуд Шеннон
(30.04.1916 – 24.02.2001)

мого любыми средствами любому получателю. В настоящее время теория Шеннона не оказывает влияния на психологические исследования в сколько-нибудь существенной степени. Тем не менее, многие понятия, унаследованные из его работ, остаются важной частью психологии обработки информации.

Влияние трансформационной лингвистики. Ноам Хомский доказывал, что язык никогда не смогут научно объяснить или понять в бихевиоральных терминах. Он настаивал на том, что этот подход коренным образом неправильно представляет себе природу языка, игнорируя его наиболее существенные свойства: структуру, правила и грамматику. С его точки зрения, язык нужно объяснять, ссылаясь на "правила в голове человека" (*rules in the head*), которые дают возможность иметь дело со структурой (т.е. системными отношениями между такими частями предложения как фразы и части сложного предложения). Для лингвиста овладение языком включает в себя усвоение (*internalizing*) системы правил, управляющих этими отношениями. Понятие структуры едва ли совместимо с бихевиористской наукой; невозможно представить себе усвоение структуры в форме научения физичиским стимулам и реакциям. Важность правил, имеющихся в голове человека, была отчетливо сформулирована лингвистами в разграничении "способность – активность" (*competence – performance*). Языковая способность является тем знанием языка, которым говорящий обладает и применяет его для того, чтобы создавать и понимать высказывания; языковая активность является реальным процессом говорения и слушания. Лингвисты и психолингвисты призывали строить теории способности, и были вынуждены постулировать существование врожденных способностей к пониманию и порождению речи. Ребенка рассматривали как "устройство овладения языком" (*language acquisition device*), предварительно запрограммированное природой на извлечение из окружающей среды той информации, которая требуется для усвоения языковой системы. В то время



Аврам Ноам Хомский
(род. 07.12.1928)

как лингвисты критиковали бихевиоральную психологию как устаревшую и бессмысленную, бихевиористы контратаковали, объявляя лингвистов антинаучными путаниками.

Обескураженная невзрачными результатами бихевиоризма в создании теорий речевого поведения, нарождающаяся парадигма обработки информации в поисках идей все больше ориентировалась на лингвистику. Система понятий и теорий Хомского вначале была заимствована во всей своей полноте.

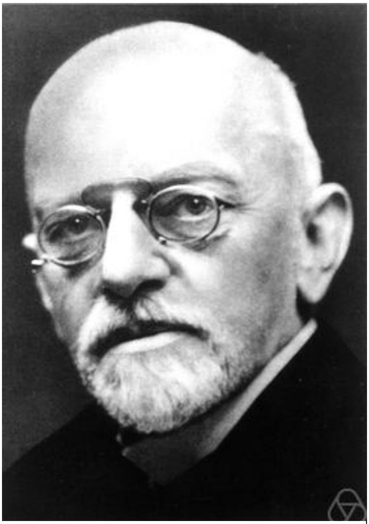
Однако, теория обработки информации теперь в меньшей степени полагается на лингвистику, чем это было некогда, разобравшись, что ряд лингвистических понятий практически бесполезен для психологии. Кроме того, лингвистика продолжила свое развитие. Значительная часть работы в психологии и лингвистике теперь сфокусирована на семантике, по-видимому, самой слабой точке в теоретической системе Хомского. Тем не менее, сохранено многое из раннего взаимодействия психологии обработки информации и лингвистики. Разграничение “способность-активность”, по-видимому, принято безоговорочно. Современные исследователи стремятся обнаружить психологические процессы или умственные операции, которые лежат в основе языковой активности. Продуктивность и креативность, отождествленные первоначально с языком, теперь с таким же успехом приписывают и др. видам когнитивной активности, включая восприятие, память, мышление и понимание.

4.1.3. Влияние теории вычислительных систем

Теория вычислительных машин и систем представляет собой семейство разнородных специальных дисциплин, включая теорию алгоритмов, численные методы, теорию автоматов, языки программирования и искусственный интеллект. Теория вычислительных систем и психология обработки информации развивались в тандеме; обе вышли из плодотворных работ по математической логике, и обе занимались природой разумного поведения. Появление вычислительных машин и концептуальных оснований, на которых они строились, способствовало возникновению еще одной метафоры для человеческих психических и интеллектуальных способностей, теоретические структуры и представляющего (*representational*) языка, с помощью которых можно было выразить теории поведения.

Ряд теорий когнитивных процессов нашел свое выражение в форме машинных моделей. Теории мышления и, в особенности, решения задач, извлекли свою выгоду из этого следования строгим формам, возможно потому, что человеческое мышление по природе своей носит сериальный (последовательный) характер.

Наиболее глубоким вкладом со стороны вычислительной техники стала альтернативная метафора для психических процессов. Подобные аналогии чрезвычайно важны в научном поиске, оказывая мощное влияние на выбор вопросов для исследования, структуру и интерпретацию экспериментов, а также на создание теории. Компьютер является предметом материальной культуры, свойства которого сравнительно хорошо понятны, и аналогия является неотразимой. Компьютеры получают входной сигнал в символической форме, перекодируют его, сопоставляют с хранящимися внутри структурами, принимают решение относительно сигнала, создают некоторые новое его выражения, сохраняет его определенную часть или весь сигнал, и выдают выходной сигнал, опять же в символической форме. По аналогии, это составляет большую часть того, чем занимается психология обработки информации: каким образом люди получают информацию, перекодируют и сохраняют ее в памяти, а затем используют для того, чтобы принимать решения и управлять доступным наблюдению поведением. Обращаясь к теории вычислительных систем, психологи, стоящие на позициях информационного подхода, строят теории человеческих способностей и поведения, используя понятия из области вычислительной техники, такие как буфер, исполняющая система, компилятор и системная архитектура.



Давид Гильберт
(23.01.1862 – 14.02.1943)

Алан Мэтисон Тьюринг
(23.06.1912 – 07.06.1954)

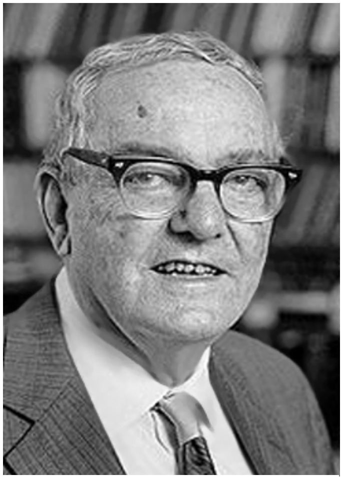
В начале XX в. Давид Гильберт бросил своим коллегам вызов, предложив формализовать интуитивные понятия доказательства, вычислимости, полноты и непротиворечивости. Алан Тьюринг описал “универсальную машину” в статье 1936 г., посвященной вопросам полноты и вычислимости. "Универсальную машину" можно представить себе как некоторую гипотетическую систему с небольшим набором основных операций, посредством которых она может решать внушительный ряд математических задач. Идеи этой работы Тьюринга предвосхитили изобретение современного цифрового компьютера; фактически, всё, что делают современные компьютеры, сводимо к фундаментальным способностям универсальной машины Тьюринга. Значимость открытия Тьюринга для когнитивной психологии заключается в том факте, что эта универсальная машина эффективно конкретизирует абстрактные процессы, связанные с обработкой символов. Абстрактные символы формальной логики можно копировать, трансформировать, переставлять в др. порядке и сочленять, так же как и физические объекты. Символы и процесс символьных преобразований становятся, тем самым, осязаемыми объектами для изучения. Этот прорыв открыл путь для демонстрации того, что, по крайней мере, некоторые человеческие идеи, умственные способности и процессы мозга можно отождествить с физическими системами символов, содержащими символические представления, которые меняются под воздействием точно определенных процессов символьных преобразований. Поэтому "ментальные события" можно описывать в теоретической системе, которая употребляется также для конкретных физических вещей.

Аллен Ньюэлл
(19.03.1927 – 19.07.1992)

Точная формулировка этого моста между абстрактным и конкретным была достижением Аллена Ньюэлла и Герберта Саймона. Гипотеза "физической системы символов" (physical symbol system) Ньюэлла и Саймона предполагает, что *важные аспекты человеческой психики, мозга и компьютера являются разными примерами систем одного и того же рода*.

Гипотеза физической системы символов, в явной или неявной форме, лежит в основе большинства исследователей и теорий в психологии обработки информации и, следовательно, должна быть экспериментально проверена и усовершенствована. По своей сути, эта концепция способна определить наличие интеллекта в системе и объяснить то, каким образом интеллектуальная система – человек (вернее – программа работы мозга), либо искусственная – компьютер (вернее – комплекс программ), создают новые знания.

На современном этапе развития человечества особенно сильное влияние на прогресс оказывает уровень развития систем управления и коммуникаций. Информационные процессы в них играют ведущую роль. Удивительно, но попытка разобраться с достижениями в области количественных оценок информации привела к неутешительным результатам. Выяснилось, что в теории информации сама «информация» не содержит смысла. Для систем управления это нонсенс. Кроме того, проектировать системы управления и управлять на основе вероятностных расчётов или предположений – явный путь к аварии или её усугублению. Поэтому изначально можно сомневаться в эффективности использования математического ожидания логарифма вероятностей сообщений (энтропии), циркулирующих в контуре управления, для количественной оценки информационных процессов в системе управления.



Герберт Александер Саймон (15.06.1916 – 09.02.2001)

Таким образом, чтобы разобраться с мерами и единицами измерения информации, необходимо начинать с понимания того, что именно измеряется.

4.2. Философское осмысление понятия «информация»

Выделено три направления исследования информационных процессов: статистическое (синтаксическое), семантическое (смысловое) и прагматическое (ценностное). Согласно этим взглядам, существует три соответствующие этим направлениям теории информации (одна из которых – теория информации К.Шеннона). Однако в работах, в том числе философских, не содержится обоснованного понятия термина «информация». Это фундаментальное понятие.

Словари толкуют информацию как сведения, а сведения как информацию.

Для однозначного толкования понятия «информация» остановимся на ряде положений как результате философского осмысления практического использования этого понятия.

Информация идеальна. По сути это смысл сведений о материальном мире, обмен которыми происходит путём «сообщений».

Сообщение – это дуальное образование из смысла и его материального носителя.

Если сообщение – форма существования информации, то способ смыслового кодирования определяет саму возможность её зарождения (генерации) и использования.

Способ смыслового кодирования F - является функцией связи между материальным носителем (кодовым обозначением) K и смысловым содержанием C :

$$F = f\left(C, K\right).$$

Таким образом, информация по природе своей триедина в том смысле, что без способа смыслового кодирования она не может быть генерирована и использована. Это существенное обстоятельство практически выпадает из поля зрения философов и негативно отражается на развитии теории информации.

На рис. 4.1 представлен процесс генерации информации как иллюстрация её триединой сущности.

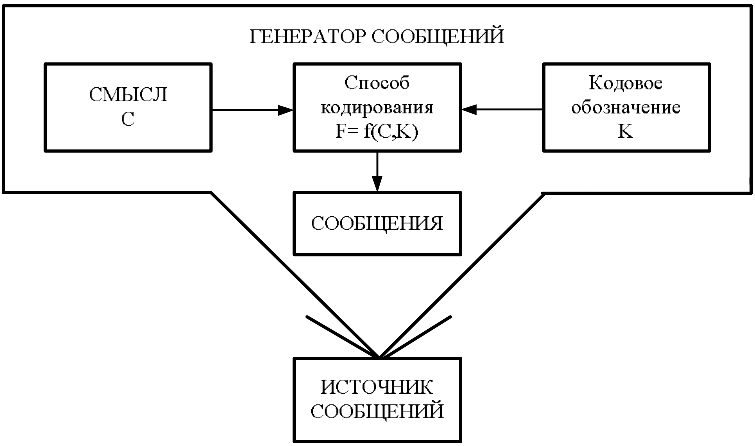


Рис. 4.1. Процесс генерации информации
(смыслового кодирования сообщений)

Исходя из изложенного, очевидно, что процесс «генерации» – это процесс преобразования материального носителя смысла из одной формы в другую. Технически это осуществляется путём, например, модуляции сигнала-носителя «информационным» сигналом. При этом имеется в виду, что смысловое содержание не меняется. Изменение смыслового содержания происходит в результате преобразования, тождественного умозаключению (в том числе при технически реализованном преобразовании).

Отражение не является всеобщим свойством материи, так как отражение – это модель «первоисточника». Для её создания и хранения должна существовать соответствующая система, которой нет, например, в неживой природе. Всеобщим свойством материи является реакция на воздействие в процессе взаимодействия объектов.

Автоматическая система не относится к объектам живой природы, однако в ней циркулирует информация, поскольку смысл реакций взаимодействующих в ней элементов заложен человеком.

Категорией, определяющей сущность понятия «информация», безусловно, является смысл. Поэтому теория информации – это теория процессов генерации, передачи, преобразования (в том числе, помехой) и использования смыслового содержания (смысла) сообщений, включая количественную оценку этих процессов. Именно в этом специфика теории информации.

Любая теория является информационной, но это не значит, что она является теорией информации. Вероятностный процесс уничтожения ракет противника в процессе атаки объектов полностью соответствующего процессу поражения помехой кодовых обозначений в линии связи, который изучал К.Шеннон.

Вероятностная теория процесса уничтожения ракет однозначно не считается теорией информации. Правомочно ли считать теорию информации К.Шеннона теорией информации только потому, что сообщения предназначены для перемещения сведений, а не взрывчатки? Действительно, скорее это просто раздел теории связи.

Совершенно очевидно, что теория информации К. Шеннона не является теорией информации в указанном выше смысле, но очевидно также, что не случайно эта теория вызвала такой ажиотаж и энтузиазм в научном мире. Не случайно, так как потребность в создании и развитии теории информации как общей «теории смысла» давно назрела.

4.2.1. Семантический аспект: усвоение переданной информации

От Приемника информация направляется к Получателю. На этом пути выполняется декодирование сигнала (процедура противоположная рассмотренному выше кодированию) и восприятие, понимание смысла сигнала Получателем.

Семантика – наука о понимании (значении) определённых знаков, последовательностей символов и других условных обозначений; верное восприятие или интерпретация какого-либо события, явления, факта. Понимание обеспечивает установление связи раскрываемых новых свойств объекта познания с уже известными субъекту, формирование операционального смысла новых свойств объекта (появление возможности ими оперировать) и определение их места и роли в структуре мыслительной деятельности. Для обозначения имеющихся сведений используется термин тезаурус.

Тезаурус – словарь языка с полной смысловой информацией, – полный, систематизированный набор данных о какой-либо области знания, позволяющий человеку или ЭВМ в ней ориентироваться, – совокупность сведений, которыми располагает пользователь.

Вместе с тем для понимания нового материала (незнакомых фактов, событий и т. д.) человек всегда должен решить определенную мыслительную задачу. С точки зрения системного подхода понимание можно определить как включение в поле зрения нового более глубокого уровня иерархии, т. е. объяснение свойств объекта взаимодействием образующих его элементов. Например, чтобы понять сезонное изменение численности мелких млекопитающих (рост летом и падение зимой), нужно исследовать процессы размножения (соотношение между половозрелыми и неполовозрелыми особями в разные сезоны года) и увидеть, что зимой нет половозрелых, а летом – есть, значит, летом идет размножение, а зимой – вымирание популяции. Понимание – это некий синтез имевшихся знаний с вновь полученными, это обогащение модели действительности новыми элементами.

Стало быть, семантические свойства информации выявляются на фоне способности пользователя принимать (понять) поступившее сообщение.

Это значит, что Получатель уже должен обладать некоторым багажом знаний в данной области (тезаурусом), а также способен к умственной работе, чтобы ассимилировать полученные сведений с имеющимися знаниями (синтез).

Проще отследить зависимость семантического наполнения информации от объема тезауруса Получателя (S_p).

Графическое представление такой зависимости приведено на рис. 4.2. Количество семантической информации I_s равно 0 при $S_p \sim 0$. когда пользователь не понимает поступающую информацию, а также при $S_p \to \infty$ когда пользователь все знает, и поступающая информация только дублирует имеющуюся, избыточна, не нужна.

Максимальное количество семантической информации I_c потребитель приобретает при согласовании ее смыслового содержания S со своим тезаурусом S_p $(S_p = S_{p\,opt})$, когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения. Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного. Для Получателей без тезауруса и способности к синтезу поступающая информация будет непонятна. Интересна только та информация, которая расширяет, увеличивает наше знание.

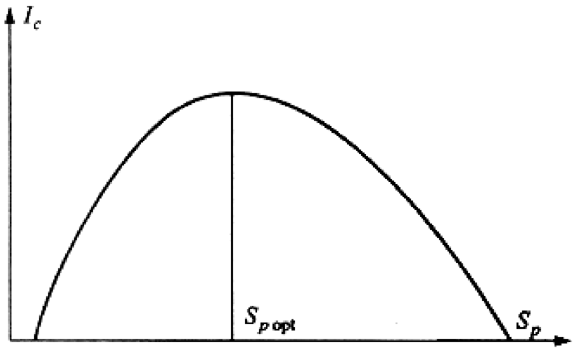


Рис. 4.2. Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса $I_c = f(S_p)$

Максимальное количество семантической информации I_c потребитель приобретает при согласовании ее смыслового содержания S со своим тезаурусом S_p $(S_p = S_{p\,opt})$, когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения. Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного. Для Получателей без тезауруса и способности к синтезу поступающая информация будет непонятна. Интересна только та информация, которая расширяет, увеличивает наше знание.

Знание можно определить как мысленный образ (мысленную модель) объекта исследования. В нем важны не только элементы знания (факты действительности), но и связи между ними (которые мы рассматриваем как причинно-следственные). Получить достоверные факты (наблюдения) о природе гораздо проще, чем объяснить их источники (построить причинно-следственные связи), т. е. построить теорию появления фактов. Создание такой модели требует большого напряжения мысли. Если получаемая информация запутана, непоследовательна, переусложнена (частично не соответствует тезаурусу Получателя), то построить на ее основе модель (образ) затруднительно. Это значит, что не только содержание, но и форма подачи во многом определяет объем усвоения информации, т. е. семантическое содержание данной информации для данного Потребителя. С этой точки зрения и дипломная работа, и статья, и доклад студента на конференции или защите должен быть тщательно продуман с той точки зрения, чтобы структура подачи материала сразу позволяла Получателям (аудитории, жюри) строить адекватную модель, объемный образ объекта исследования. Такое же требование важно предъявлять и к иллюстрациям, например, к картам. Пренебрежение к пользователю у составителей карт, по-видимому, традиционно потому, что сама карта отнимает огромные силы. Желание вложить в ограниченный лист бумаги максимум информации приводило и приводит к заведомой избыточности, требующей выработки специальных навыков чтения, которые доступны только изощренным специалистам. Это одна из причин, почему в нашей стране практически не используется огромное количество тщательно созданные и весьма информативных региональных атласов, в составлении которых наша картография достигла совершенства.

Итак, понимание сообщения = синтез (тезаурус + информация). … *В идеале необходимо ... чтобы информация обращалась к языку образов, которыми в норме оперирует человек и соответственно генерировала бы самые разнообразные и самые смелые ассоциации со сферой ее деятельности ... Необходимо, чтобы в ходе восприятия возник синергетический эффект. Эта задача сродни искусству, но требования совершенно другие. ... Все люди одного и того же статуса должны извлекать достаточно сходную информацию вне зависимости от места и времени... сродни задачам, решаемым создателями плакатов.*

Все сказанное выше относилось к обмену информацией между людьми. Для исследовательской деятельности также подходит схема информационного канала связи. В этом случае отправителем информации выступает природа. Стоит указать на тот момент, что для расшифровки «сообщений природы» существует обширный аппарат статистики, методов линейной алгебры, различных методов многомерного анализа. К сожалению, в этом случае «Отправитель» не обращает внимания ни на тезаурус Получателя, ни на его способность строить адекватные модели этих сообщений. Можно лишь перечислить некоторые вопросы, ответы на которые помогают исследователю осмыслить получаемую информацию:

1. Как из множества наблюдений выделить переменные?
2. Как отобразить эти переменные в собственном алфавите с минимальными искажениями?
3. Как выявить отношения этих переменных друг к другу?
4. Как установить порядок в этих отношениях?Как свести множество сочетаний состояний к элементарным непротиворечивым формулам и высказываниям?
5. Как объяснить (доказать), почему существую именно такие элементарные высказывания?

По существу эти вопросы есть фрагмент практического использования системного подхода, они относятся к принципам системности, структурно-функциональной организации, иерархичности, эмерджентности . Не хватает вопросов, относящихся к еще трем принципам, однако главное явно – в точки зрения теории передачи сигнала исследование природы есть декодирование и понимание «смысла» (моделирование) информации, заключенной в природе.

4.2.2. Прагматический аспект: ценность информации для получателя

Прагматическое содержание полученной информации определяется тем, помогла ли она Получателю в продвижении к цели?

Стремясь определить ценность информации относительно целей Получателя, следует определить, что такое есть цель.

Цель не есть отражение действительности такой, как она есть, а такой, как она должна стать в результате деятельности. Цель – то, что представляется в сознании и ожидается в результате определенным образом направленных действий. Цель – это образ желательного результата деятельности.

Чем более способствует информация достижению цели, тем более она ценна.

Информация выступает как объективный фактор, как носитель ценности, однако субъект является субъективным фактором ценности. Дело в том, что цели своей деятельности не приходят сами откуда-то со стороны, но формируются и формулируются самим исследователем. Цели получают в результате некоего (пусть даже и мысленного) «моделирования» системы отношений «объект–окружающая среда». Если модель собственной деятельности, построенная Получателем информации, ложна, неадекватно отражает окружающий мир, то могут быть ложны и цели. В этом случае полученная объективная информация, реально отражающая внешний мир, но противоречащая ложным целям, приобретает отрицательную ценность или вообще отбрасывается субъектом (нулевая ценность).

Этот круг можно разорвать, если рассматривать не субъективное состояние субъекта относительно целей, порожденным его собственным видением мира, а объективное его состояние относительно реальных условий его жизнедеятельности. Ценность информации может быть принято эквивалентной расстоянию субъекта от границы области устойчивости. Ценность такой информации тем выше, чем более надежно она предупреждает субъекта о приближающейся границе устойчивости.

Будучи сторонниками широкого обмена информацией, мы должны постоянно иметь в виду, что избыток информации столь же вреден, как и ее недостаток, что любая информация полезна лишь тогда, когда она в количественном, и в семантическом, и в прагматическом плане сопоставима с возможностями всех функциональных блоков информационной системы, так что любое «сообщение» с необходимой для поставленных целей скоростью проходит от системы измерения к системе принятия решений, способной с семантической и технических точек зрения принять адекватное решение. Кроме того, следует отметить, что информация есть атрибут взаимодействия.

Будучи порождением чисто материальных отношений, она сама по себе не материальна – не материальна мысль, переданная материальными агентами.

На информацию не распространяется физические законы сохранения, другое дело, что пока не разработаны объективные меры семантического и прагматического объемов информации. Так, по мере использования информации об окружающем мире для самоорганизации ценность ее снижается. В таких соотношениях можно усмотреть проявление законов сохранения: количество некоторой информации есть постоянная величина; информация из окружающей среды перешла в систему, исчерпана из окружающей среды и соответственно потеряла свою ценность, а среда по соответствующей переменной стала полностью предсказуемой и управляемой. Но возможно ли такое, например, для прогноза погоды?

4.3. Терминология теории информации

Исходя из анализа процессов генерации информации, её передачи, преобразования, хранения и др., можно видеть, что понятию «информация» придаётся разный смысл. Возникает необходимость разобраться в терминологии. В дальнейшем основополагающим в терминологии теории информации положением будем считать триединую природу информации. Все указанные процессы – это процессы со смыслом информации. Поэтому естественно и допустимо отождествление терминов «информация» и «смысл».

На практике невозможно воспроизведение смысла без способа его кодирования, однако это не мешает его существованию, например, в тексте, памяти ЭВМ и др. памяти. Поэтому возникает необходимость введения специального термина – сообщение, как совокупности (дуальности) смысла и его материального носителя. Наконец, можно изучать процессы с самим материальным носителем смысла – «сигналами», кодовыми обозначениями, «тарой» для хранения и транспортировки смысла.

Очевидно, что теория информации – это теория смысла. Поэтому выделение её разделов: статистическая теория, семантическая теория и прагматическая теория информации – на данном этапе её развития уже не представляется продуктивным.

На практике используется также термин «сведения». В свете изложенного это синоним термина «информация». Однако, и словосочетание «смысл сведений» не является тавтологией, когда речь идёт об умозаключениях.

Необходимо различать в сообщениях смысл вообще (измеряемый) и смысл содержательный, предметный, в частности, используемый при решении оператором конкретной задачи управления. В автоматических системах управления, как и в системах передачи сообщений, содержательная информация присутствует опосредованно. Поэтому при оценке, например, эффективности обработки информации (автоматическими) техническими устройствами, работающими с кодовыми обозначениями смысла, правомочно условиться говорить о «емкостной» информации как содержательной, предметной информации. Тогда понятно, что информация о количестве искаженных помехой символов в кодовом обозначении – это предметная, содержательная информация, связанная с теорией систем передачи сообщений. Это «емкостная» теория о минимизации длины кодового обозначения, способах восстановления искажённых символов (помехоустойчивое кодирование) и т.п. Смысл этой теории - о ёмкости, а не о смысле как таковом.

4.3.1. Характерные неудачные попытки введения меры и единицы измерения количества информации

Введение К. Шенноном меры и единицы измерения количества информации можно считать неудачной попыткой как в части определения самого понятии информации, обоснованности ввода меры и единицы её измерения, так и в части того, что именно и как именно оценивается в каналах связи. Это внесло серьёзную путаницу в терминологию теории информации и, что особенно вредно, было на веру воспринято и растиражировано математиками. Вот характерный пример.

Пример 4.1. Допустим, что нам интересно знать, сдал или не сдал экзамен данный студент. Примем следующие вероятности этих двух событий:

$$P(\text{сдал}) = 7/8\,,\ P(\text{не сдал}) = 1 - P(\text{сдал}) = 1/8\,.$$

Отсюда видно, что этот студент является довольно сильным. Если нам сообщили, что он сдал экзамен, мы вправе сказать. «Ваше сообщение мне мало что дало, я и без этого предполагал, что он сдал». Количественно информация этого сообщения равна:

$$I(\text{сдал}) = \log_2(7/8) = 0{,}193\ \text{бита.}$$

Если нам сообщили, что не сдал, мы скажем: - «Неужели?» и почувствуем, что в большей степени обогатились знаниями. Количество информации такого сообщения равно:

$$I(\text{не сдал}) = \log_2(1/8) = 3\,.$$

Вот еще одна иллюстрация ограниченности подхода теории информации Шеннона: «Пусть на одном листе напечатан смысловой текст на русском языке, на другом – хаотическая последовательность букв, отпечатанная, скажем, обезьяной. Первый текст имеет энтропию H(X) = = 2 бит/симв, тогда как второй, составленный из равновероятных и независимых символов, имеет максимальную энтропию $H_{\max}(X) = 5$ бит/симв. Приходим к выводу, что обезьяний текст более информативен, чем осмысленный. Этот парадокс объяснить несложно. Попробуйте прочитать и заучить такой текст и затем сравните свои усилия с чтением наизусть стихов Пушкина. Эти усилия как раз и пропорциональны энтропии как степени неопределенности букв текста. Что же касается ценности обезьяньего текста, то он годится лишь для оценки её похвальных (и порой успешных) усилий подражать человеку».

Можно видеть, что рассматривается опыт с двумя исходами и энтропией (канал связи без помех):

$$H = -\sum_{i=1}^{n} p_i \log p_i\ .$$

Следовательно, количество информации, приходящей с сообщением в этом канале $I = H = 0{,}544$ бит.

Действительно, $\log_2(7/8) = 0{,}193$, а $\log_2(1/8) = 3$. А вот вопрос о том, почему автор, получив второе сообщение из области да - нет, почувствовал, что в большей степени «обогатился знаниями», оставим на совести автора примера.

Если этот подход к оценке количества семантической информации обсуждать серьёзно, то очевидно следующее. В качестве меры количества информации К. Шеннон предложил математическое ожидание случайной величины – $\log_2 p_i$, т.е. энтропию опыта со случайными исходами $H = -\sum_i p_i \log p_i$, характеризующий источник информации в целом.

Логарифм вероятности любого исхода сдачи экзамена студентом однозначно связан с этой вероятностью. Поэтому эмоции вызывает в первую очередь вероятность сама по себе. К количественной оценке информации, предложенной К. Шенноном, она отношения не имеет.

Этот пример и эта убеждённость, что именно так предложил измерять количество информации в сообщении К. Шеннон, постоянно тиражируется.

Пример 4.2 (о количестве смысловой информации). Принята следующая количественная мера информации. Пусть A и B – события некоторого вероятностного пространства. Количество информации $I(A/B)$, которое заключается в событии (сообщении) B относительно A, определяется как число:

$$I(A/B) = \log \frac{P(A/B)}{P(A)}.$$

Появление события $B = A$ можно интерпретировать как сообщение о том, что наступило A. Тогда $I(A/A) = I(A)$ есть по определению количество информации $I(A)$, заключающейся в сообщении A:

$$I(A/A) = I(A) = -\log P(A).$$

Таким образом, ясно, что понимание сущности информации и вычисления количества информации в сообщениях, продемонстрированное в приведённых примерах, никак не связано с предложенной К. Шенноном количественной мерой информации.

Понятно, что математику никогда не придёт в голову, что случайная величина и её математическое ожидание имеют несовместимую физическую природу. Однако, (по К.Шеннону) это так, и это правильно по определению. В приведённом примере со студентом считается, что одно из сообщений содержит 3 бита информации и, в принципе, может быть сколь угодно большим. Понятно, что это нонсенс, так как $H_{\max}$, «тара» (по К.Шеннону), для сообщения рассматриваемого источника не может превышать 1 бит.

Что касается неосознанного введения авторами примеров количественной меры семантической информации, то причина этой их ошибки кроется в том, что смысловое кодирование вообще не является вероятностным процессом.

4.3.2. Линейная мера количества семантической информации

Смысловое соответствие сведений и их кодовых обозначений при «измерениях смысла» будем считать известным. Вопрос в том, что именно будет измеряться и какой мерой. В самом общем случае реально измерить «многое, понимаемое как единое целое», т.е. (по определению) множество, и ничто другое. Измерить «многое» означает определить величину мощности множества. Смысл «единицы» измерения совпадает со смыслом признака «единого целого» в каждой «реализацией опыта».

Очевидно, что одной мерой (единицей измерения) нельзя мерить, например, расстояние и вес. Чтобы измерять общее количество смысла того или иного источника, смысл сведений должен быть однотипным. Тогда при выбранной единице измерения полученное при измерении число будет тем больше, чем больше смысла заключено в источнике.

При (семантическом) кодировании выполняется требование: элементы одного множества должны иметь нечто общее по смыслу и при этом различаться. В этом случае можно говорить о «порциях смысла» и их общем числе. Назовём эту каждую «порцию смысла» сенсом (sense). Сенс – это собственно смысл каждого кодового обозначения в некотором их множестве, генерируемом при кодировании смысла источника сообщений. Это то, общее количество чего измеряется в источнике.

Используем линейную меру для измерения. В качестве единицы измерения выберем 1 (сенс). Тогда, если источник содержит n элементов, то он содержит n (сенс) смысла.

Пример 4.3. Закодируем четырьмя разными цветами смысл следующих возможных состояний («многого») объекта («единого целого»): повреждён, выведен из действия, готов к работе, функционирует по прямому назначению. Можно сказать, что общее количество смысла в источнике равно 4 (сенсам).

Таким образом, введённые линейная мера измерения количества (смысла) информации и её единица – (сенс) – действительно позволяют измерять количество (общий объём) смысла в источнике (приёмнике) информации.

С другой стороны, если общий объём смысла в источнике составляет n (сенс), то единица измерения (1 сенс) как реализация опыта содержит $1/n$ общего количества смысла в источнике (n сенс).

4.3.3. Логарифмическая мера как мера количества ёмкости

Рассмотрим физический смысл выражения $1 = \log_N N$. Оно прекрасно демонстрирует связь единиц линейной и логарифмической мер. Суть этого выражения в том, что единица количества информации в одной реализации (единица) однозначно связано логарифмической мерой с общей информационной мощностью источника (N), измеренной линейной мерой с единицей (сенс). Создаётся впечатление, что достаточно придумать название этой смысловой единице, однозначно связанной элементарной математической (логарифмической) функцией с линейной мерой, и проблема измерения количества смысловой информации решена. А именно: пользуйся мерой, которая удобна, так как обе подходят. На самом деле это заблуждение.

В действительности если N (сенс) – это количество смысла в источнике информации, то $\log_N N$ (и по любому другому основанию) теряет смысловое содержание, поскольку теряет смысл функция кодирования $F = f(C,K)$.

Таким образом, **линейная мера – единственная мера для измерения количества смысла**. Единицей измерения количества порций смысла в источнике является (сенс). А логарифм количества смысла – это (из-за отсутствия способа кодирования) полная бессмыслица.

Физический смысл единицы логарифмической меры – это **ёмкость «тары»** для транспортировки и хранения смысла. Её тоже нужно измерять. Информация о ёмкости тары – это «емкостная» информация.

Допустим, что источников информации несколько, например, n. Все они характеризуют состояние более сложного объекта. Общее число элементов в них равно $K = \sum_i n_1 + n_2 + ... + n_n$.

Число состояний сложного объекта будет характеризоваться уже не числом K, а числом:

$$N = \prod_i n_i \ , \ i = 1, 2, ..., n .$$

Ёмкость этого источника:

$$E_0 = \log N = \sum_i \log n_i = \prod_i n_i .$$

Выражение $\sum_i \log n_i = \prod_i n_i$, можно трактовать как закон сохранения количества емкостной информации ($E_0 = \log N$): общее количество емкостной информации не зависит от способа её распределения по «емкостям» для хранения и транспортировки смысловой информации ($\sum_i \log n_i = \prod_i n_i$).

4.3.4. Универсальная единица ёмкости источника сообщений

В качестве универсальной единицы ёмкости материального носителя логично использовать $\log_2$, так как это ёмкость для минимально мощного источника из двух сообщений. Меньше не бывает, так как исчезает возможность выбора. В ситуации, когда может произойти только одно единственное событие, необходимо иметь в виду, что кодируется два сообщения: событие произошло и не произошло.

Название этой единицы измерения емкости (материального носителя) для транспортировки и хранения смысла общеизвестно:

$$1 = \log_2 2 = 1\,(\text{бит}).$$

4.3.5. Ёмкость источника сообщений и энтропия равномерного распределения

Величина $1/n$, понимаемая исключительно как доля количества общего смысла источника, измеренного линейной мерой и приходящегося на сообщение (1 сенс), однозначно характеризует смысловую мощность источника (через n). Но каждой порции смысла соответствует материальный носитель.

Ёмкость источника в связи с этим является величиной детерминированной и измеряется логарифмической мерой.

Величина $\log 1/n = -\log n$. Но модули этих величин равны: $|\log n| = |-\log n|$. Следовательно, введение и таким образом логарифмической меры приводит по существу к одному результату.

Если $1/n$ трактовать как величину вероятности равномерного распределения, то в этом случае ёмкость материального носителя определялась бы путём вычисления энтропии равномерного распределения этих вероятностей:

$$H = -\sum_{i=1}^{n} p_i \log p_i \; .$$

В связи с тем, что ёмкость источника – величина детерминированная, правильно будет говорить о простом совпадении численных результатов вычислений.

Равномерное распределение вероятностей является своеобразной гранью, характеризующей (а потому присущей) как вероятностный, так и детерминированный процессы. Поэтому теоретически такому распределению присущи свойства обоих процессов. Но величина $1/n = p_i$ могла бы быть истолкована как равномерное распределение вероятностей n событий только в случае, если процесс кодирования носил бы вероятностный характер.

Уместно вернуться к трактовке энтропии H (как количеству емкостной информации). Легко видеть, что энтропия произвольного распределения (отличного от равномерного) как количественная мера емкостной информации искажает представление о мощности собственно источника сообщений из-за наложения случайного процесса передачи.

Обратим внимание на то, что вопрос о количественной оценке информационной мощности некоторого источника (приёмника) информации можно исследовать как с целью выяснения величины интегральной смысловой мощности собственно источника, так и в части количества информации, приходящегося на сообщение, однозначно связанного с возможностями выбора, т.е. мощностью источника.

Эти вопросы связаны, поскольку реализация передачи или приёма одного конкретного сообщения (величиной 1сенс) одновременно несёт информацию о том, что $n-1$ сообщение (источника мощностью n сенс) не получило реализации. Весомость этой информации тем больше, чем больше величина n. Следовательно, конкретная реализация каждого сообщения всегда связана со всей смысловой мощностью источника. Эта связь проявляется в величине потребной ёмкости кодового обозначения, измеряемой логарифмической мерой.

Таким образом, показано, что единственной мерой измерения смысла является линейная мера, которой измеряется количество «порций смысла» (мощность источника смысла). Единицей измерения количества смысла является 1 (сенс).

Логарифмическая мера используется для измерения ёмкости материального носителя. Универсальной единицей измерения ёмкости «емкостной информации» является (бит).

4.4. Основы классической теории информации

Начальным этапом классической теории информации является анализ информационных характеристик источников. В силу различия форм, которые принимают сообщения, различаются источники дискретных и непрерывных сообщений. Отличительной особенностью дискретных источников является то, что их алфавит X состоит из конечного набора символов x_i, т.е.

$$X = \{ x_1 ,\ x_2 ,\ldots,\ x_N \}, \tag{4.1}$$

где N – объём алфавита (число символов источника). Примеры дискретных источников: клавиатура компьютера с набором букв, цифр, знаков препинания в качестве символов; цифровые часы и пр. Статистические свойства дискретных источников описываются вероятностями $P(x_i)$ появления символов x_i.

Источники непрерывных сообщений вырабатывают сообщения с бесконечным (континуальным) числом состояний. Обозначим множество возможных состояний источника как сегмент:

$$X = \left[x_{\min}, \mathrm{x}_{\max} \right], \tag{4.2}$$

при этом $x \in X$ рассматривается как непрерывная случайная величина с плотностью вероятности $p(x)$, либо как случайный процесс $x(t)$ с одномерным распределением $p(x,t)$ (при наличии зависимости от времени).

4.4.1. Информационные характеристики дискретных источников

Рассмотрим источник дискретных символов из алфавита (4.1) с априорными вероятностями появления символов $P(x_i)$, $i = 1, 2, ..., N$. Вероятности $P(x_i)$ характеризуют относительную частоту появления символов. Например, в русском языке наиболее часто встречается буква «О», имевшая вероятность $P(O) = 0{,}11$, наиболее редко – буква «Ф» с вероятностью $P(\Phi) = 0{,}002$. Это значит, что в какой-нибудь толстой книге в каждой 1000 букв в среднем можно встретить 110 букв «О» и 2 буквы «Ф». Очевидно, что для полной группы событий x_i выполняется условие:

$$\sum_{i=1}^{N} P(x_i) = 1 . \tag{4.3}$$

Если в последовательности символов источника все символы независимы, и условная вероятность:

$$P(x_i / x_k) = P(x_i),\ i,k = 1, 2, ..., N, \tag{4.4}$$

то говорят, что источник не обладает памятью.

При наличии взаимной корреляции между двумя символами

$$P(x_i / x_k) \neq P(x_i), \qquad (4.5)$$

т.е. источник обладает памятью на один символ. Более сложные источники имеют память на 2, 3 и т.д. символа.

4.4.2. Количество информации в элементе сообщения (символе)

Сравним два сообщения:

А. Завтра погода будет отвечать прогнозу: тепло и солнечно.

Б. Завтра вопреки прогнозу будет ураган.

Если первое сообщение несёт мало информации, то второе, маловероятное, сенсационное, содержит новизну и, следовательно, достаточно информативно. С точки зрения здравого смысла естественно ввести такую количественную меру информации $I(x_i)$ в символе сообщения x_i, для которого бы выполнялись условия:

1. Чем маловероятней сообщение x_i, т.е. чем меньше $P\left(x_i\right)$, тем больше в нём количества информации $I(x_i)$;

2. $I(x_i) = 0$ при $P\left(x_i\right) = 1$ – известное сообщение информации не содержит;

3. Для независимых сообщений x_i, x_k должно выполнятся свойство аддитивности:

$$I(x_i, x_k) = I(x_i) + I(x_k).$$

Можно показать, что всем трём условиям удовлетворяет лишь логарифмическая функция:

$$I(x_i) = \log_m \frac{1}{P(x_i)}.$$

Основание логарифма m характеризует единицу измерения информации. В дальнейшем примем наиболее распространённую двоичную единицу информации – бит ($m = 2$), при этом:

$$I(x_i) = \log_2 \frac{1}{P(x_i)} = -\log_2 P(x_i) \text{ [бит]}. \qquad (4.6)$$

Термин «бит» происходит от начальных и конечной букв английских слов binary unit – двоичная единица.

Для упрощения записи основание логарифма 2 ниже будем опускать.

Например, для равновероятного русского алфавита с $N = 32$:

$$P(x_i) = \frac{1}{N} = \frac{1}{32} \Rightarrow I(x_i) = \log 32 = 5 \text{ бит.}$$

Это значит, что каждую букву алфавита можно закодировать пятью двоичными символами 0 и 1, к примеру:

$$\text{А} \to 01101; \quad \text{В} \to 11010.$$

Количество информации в последовательности из 2-х независимых символов, для которых справедливо $P(x_i, x_k) = P(x_i)P(x_k)$ определяется:

$$I(x_i, x_k) = \log \frac{1}{P(x_i)P(x_k)} = \log \frac{1}{P(x_i)} + \log \frac{1}{P(x_k)} = I(x_i) + I(x_k). \quad (4.7)$$

Не представляет труда обобщение расчёта на последовательность из произвольного числа символов.

Мерой $I(x_i)$ неудобно пользоваться при вычислении количества информации, заключённого в больших массивах символов (скажем, в файле или книге). Для этого используют усреднённую характеристику – энтропию.

4.4.3. Энтропия дискретного источника без памяти

Энтропия есть среднее количество информации, содержащееся в одном символе источника. Она определяется как математическое ожидание количества информации совокупности случайных величин $I(x_i)$

$$H(X) = \sum_{i=1}^{N} P(x_i) \cdot I(x_i) = \sum_{i=1}^{N} P(x_i) \cdot \log \frac{1}{P(x_i)} = -\sum_{i=1}^{N} P(x_i) \cdot \log P(x_i). \quad (4.8)$$

Энтропия $H(X)$ характеризует источник X в целом, а не его отдельные символы. По существу, она является мерой средней неопределённости одного символа дискретного источника.

Пример 4.4. При подведении итогов успеваемости за учебный год было подсчитано, что из тысячи полученных оценок оказалось:

x_1 – "отлично"	125 оценок
x_2 – "хорошо"	250 оценок
x_3 – "удовлетворительно"	500 оценок
x_4 – "неудовлетворительно"	125 оценок

Количество информации в каждом из событий x_i равно:

$$I(x_3) = \log \frac{1}{P(x_3)} = \log 2 = 1 \text{ бит},$$

$$I(x_2) = \log \frac{1}{P(x_2)} = \log 4 = 2 \text{ бит},$$

$$I(x_1) = I(x_4) = \log 8 = 3 \text{ бит}.$$

Более редкими и, соответственно, наиболее информативными оказались события x_1 и x_4 (действительно, отличные знания, равно как и их отсутствие, выявляются у сравнительно малой части студентов). Однако энтропия как средняя характеристика, находящаяся в пределах от 1 до 3 бит на символ, будет ближе к $I_{\min} = 1$ бит для наиболее частых событий x_3. В самом деле, согласно (4.8):

$$H(X) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + 2 \cdot \frac{1}{8} \cdot 3 = 1.75 \frac{\textit{бит}}{\textit{симв}}.$$

Заметим, что аналогичным усреднением подсчитывается средний бал

$$m(X) = \sum_{i=1}^{4} x_i P(x_i) = 5 \cdot \frac{1}{8} + 4 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 2 \cdot \frac{1}{8} =$$

$$= \frac{5 \cdot 125 + 4 \cdot 250 + 3 \cdot 500 + 2 \cdot 125}{1000} = 3.375.$$

Нетрудно понять, что с ростом объема алфавита N энтропия как степень информативности (и неопределённости) одного символа возрастает. Это становится очевидным, если сравнить информативность буквы русского языка с информативностью одного иероглифа японского (или китайского). Число таких иероглифов достигает тысяч, вместе с тем каждый иероглиф может быть равнозначен слову или понятию.

Термин «энтропия» в теории информации заимствован из термодинамики, где этим понятием характеризуется степень неопределённости состояния молекул вещества, их хаотичность.

Рассмотрим на примере, как изменяется энтропия в зависимости от переопределения вероятностей $P(x_i)$.

Пример 4.5. Пусть источник имеет лишь два состояния: $X = \{0,1\}$ с априорными вероятностями $P(0) = p$, $P(1) = 1 - p$. Энтропия источника:

$$H(\mathrm{X}) = p \cdot \log \frac{1}{p} + (1-p) \cdot \log \frac{1}{1-p}\,. \qquad (4.9)$$

При изменении p в возможных пределах от 0 до 1 в крайних точках этого интервала возникает неопределённость типа $0{\cdot}\infty$. Определим предел

$$\lim_{p\to 0} \mathrm{H}(\mathrm{X}) = -\lim_{p\to 0}(p \log p) - \lim_{p\to 0}(1-p)\log(1-p) =$$

$$= -\lim_{p\to 0} \frac{(\log p)'}{\left(\frac{1}{p}\right)'} = \lim_{p\to 0} \frac{\frac{1}{p \ln 2}}{\frac{1}{p^2}} = \lim_{p\to 0} \frac{p}{\ln 2} = 0.$$

При определении предела использовалось правило Лопиталя. При $p \to 1$ получим тот же результат, заменяя $1 - p = q$ и принимая $q \to 0$.

Экстремум функции (1.9) находим равенства:

$$\frac{d}{dp} H(X) = -\log p - p\frac{1}{p \ln 2} - (-1)\log(1-p) + \frac{1-p}{(1-p)\ln 2} = 0$$

$$\Rightarrow \log p = \log(1-p) \Rightarrow p = \frac{1}{2}\,.$$

В этой точке $H_{\max}(X) = 2 \cdot \frac{1}{2} \log 2 = 1 \frac{бит}{симв}$.

График зависимости (4.9) приведён на рис.4.3.

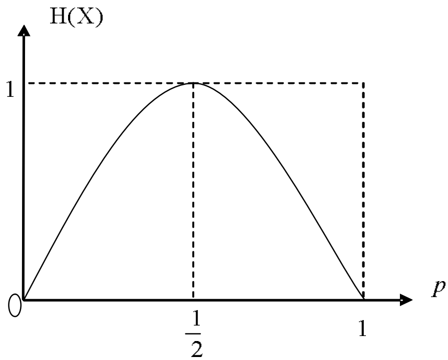


Рис. 4.3. Зависимость энтропии
от вероятности использования символа

На этом частном примере показано, что энтропия максимальна для равновероятных символов $P(1) = P(0) = 0,5$. Действительно, в этом случае нельзя отдать предпочтения ни одному из символов, степень неопределенности исхода опыта максимальна, равно как и информативность каждого символа и источника в целом. Именно этим обстоятельством обусловлено то, что при проектировании двоичных (бинарных) каналов связи стремятся обеспечить равновероятность передачи 1 и 0.

Сделанный вывод оказывается справедливым и для алфавитов с произвольным числом символов N. В этом случае энтропия максимальна для равновероятных символов с вероятностями $P(x_i) = 1 / \mathrm{N}$ и равна:

$$H_{\max}(X) = \sum^{N} \frac{1}{N} \log N = \log N . \qquad (4.10)$$

Заметим, что мера (4.10) была введена ещё в 1928 г. американским учёным Р.Хартли и носит название меры Хартли. Она является частным случаем меры Шеннона (4.8).

Ральф Винтон Лайон Хартли
(30.11.1888 – 01.05.1970)

С использованием среднего количества информации на символ $H(X)$ суммарное количество информации, заключённое в последовательности из n независимых символов, равно

$$I_n(X) = n \cdot H(X) \text{ [бит]}. \qquad (4.11)$$

4.4.4. Энтропия источника с памятью

В книге Жюля Верна “Дети капитана Гранта” спасателям потерпевшего бедствие капитана пришлось при расшифровке найденной в бутылке записки дополнять её частично смытый водой текст. Возможность такого восстановления текста во многом определяется корреляцией (зависимостью) между буквами и словами. Из-за неоднозначности восстановления текста район поиска героям книги трижды пришлось менять, однако итог его оказался успешным.

Очевидно, наличие зависимости между элементами сообщения снижает энтропию как степень неопределённости символов.

Рассмотрим источник с памятью на один символ.

Пусть: $X = \{x_1, x_2, ..., x_N\}$ – алфавит предшествующих символов, $X' = \{x'_1, x'_2, ..., x'_N\}$ – алфавит последующих символов, $P(x'_k / x_i)$ – условная вероятность появления символа x'_k после символа x_i.

Ситуация, соответствующая последовательной передаче двух произвольных символов x_i, x'_k, наглядно представлена с помощью графа передачи на рис.4.4.

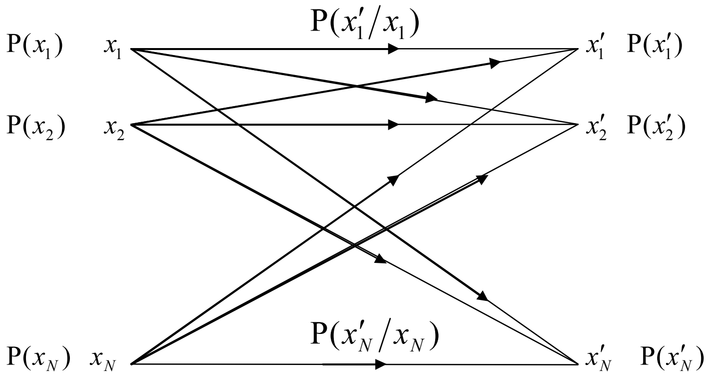


Рис.4.4. Граф передачи

Вероятность последовательного появления символов x_i, x'_k равна

$$P(x_i, x'_k) = P(x_i) \cdot P(x'_k / x_i) . \tag{4.12}$$

Введём частичную энтропию: $H(x_i, x'_k) = P(x_i, x'_k) \cdot \log \dfrac{1}{P(x_i, x'_k)}$.

Полная энтропия для алфавитов X, X' определяется суммированием по всем индексам частичных энтропий:

$$H(X, X') = \sum_{i=1}^{N} \sum_{k-1}^{N} P(x_i, x'_k) \cdot \log \frac{1}{P(x_i, x'_k)} .$$

С учётом (4.12)

$$H(X, X') = \sum_{i,k=1}^{N} P(x_i) \cdot P(x'_k / x_i) \cdot \log \frac{1}{P(x_i) \cdot P(x'_k / x_i)} =$$

$$= \sum_{i=1}^{N} P(x_i) \cdot \log \frac{1}{P(x_i)} \sum_{k=1}^{N} P(x'_k / x_i) + \sum_{i,k=1}^{N} P(x_i) \cdot P(x'_k / x_i) \cdot \log \frac{1}{P(x'_k / x_i)}.$$

В этом выражении:

$\sum_{k=1}^{N} P(x'_k / x_i) = 1$ согласно условию нормировки,

$\sum_{i=1}^{N} P(x_i) \cdot \log \frac{1}{P(x_i)} = H(X)$ – энтропия символов из X,

$$\sum_{i,k=1}^{N} P(x_i) \cdot P(x'_k / x_i) \cdot \log \frac{1}{P(x'_k / x_i)} = H(X'/X) \quad – \tag{4.13}$$

условная энтропия появления символов из X' после X.

Итак,

$$H(X, X') = H(X) + H(X'/X). \tag{4.14}$$

Таким образом, среднее количество информации, которое переносят два соседних символа из X, X', равно сумме энтропий первого символа и условной энтропии второго при условии появления первого.

Отсюда условная энтропия $H(X'/X) = H(X, X') - H(X)$ – это среднее количество информации, которое приносит последующий символ при условии, что предыдущий уже известен (известная информация $H(X)$ вычитается). Порядок усреднений при построении условной энтропии определяется схемой

$$H(x'_k / x_i) = P(x'_k / x_i) \cdot \log \frac{1}{P(x'_k / x_i)} \Rightarrow H(X'/x_i) = \sum_{k=1}^{N} H(x'_k / x_i) \Rightarrow$$

$$\Rightarrow H(X'/X) = \sum_{i=1}^{N} P(x_i) \cdot H(X'/x_i) = \sum_{i,k=1}^{N} P(x_i) \cdot P(x'_k / x_i) \cdot \log \frac{1}{P(x'_k / x_i)}.$$

Условная энтропия обладает следующим свойством:

$$H(X'/X) \le H(X'), \tag{4.15}$$

т.е. наличие корреляции между символами снижает энтропию как меру неопределённости ($H(X')$ – безусловная энтропия символов из X').

Равенство в (4.15) достигается лишь при независимых символах из X и X'. При этом, как следует из (4.14), энтропия

$$H(X, X') = H(X) + H(X') \tag{4.16}$$

максимальна. Последнее равенство характеризует свойство аддитивности количества информации для двух независимых символов из X, X'.

При определении безусловной энтропии $H(X')$ следует по формуле полной вероятности рассчитать безусловные вероятности

$$P(x'_k) = \sum_{i=1}^{N} P(x_i) \cdot P(x'_k / x_i). \tag{4.17}$$

Пример 4.6. Рассмотрим алфавит русского языка и примем $N = 32$. Возможны следующие случаи:

1. Все буквы равновероятны независимы.

Согласно (4.10)

$$H_{\max}(X) = \log 32 = 5\,бит/симв. \tag{4.18, а}$$

2. Буквы алфавита независимы и неравновероятны.

Пользуясь таблицей вероятностей букв русского языка получим:

$$H(X) = 4.39\,бит/симв.$$

3. С учётом корреляции между двумя буквами

$$H(X) = 3.52\,бит/симв.$$

4. С учётом корреляции между всеми буквами

$$H(X) \approx 2 \text{ бит/симв.} \tag{4.18, б}$$

Исходя из изложенного, сформулируем свойства энтропии $H(X)$.

4.4.5. Свойства энтропии

1. Энтропия есть вещественная неотрицательная величина: $H(X) \geq 0$.

Это свойство обусловлено тем, что вероятности $P(x_i) \in [0, 1]$, при этом каждый член суммы (4.8) неотрицателен.

2. Энтропия известного сообщения равна 0. Действительно, из:

$$P(x_i) = 1 \Rightarrow P(x_k) = 0, \quad k = 1, 2, ..., N, \quad k \neq i,$$

$$H(X) = \sum_i P(x_i) \cdot \log \frac{1}{P(x_i)} = 0 \qquad .$$

3. Энтропия максимальна и равна $H_{\max}(X) = \log N$ для независимых равновероятных символов.

Следствие: энтропия растет с ростом объёма алфавита N источника.

4. Энтропия снижается при:

– неравновероятности символов источника,

– наличии корреляции между символами.

Пример 4.7. Пусть двоичный источник с памятью на один символ вырабатывает последовательно символы из $X = \{1,0\}$ и $X' = \{1',0'\}$ с вероятностями:

$$P(1) = \frac{3}{4};\ \ P(0) = \frac{1}{4};\ \ P(1'/1) = \frac{7}{8};\ P(0'/1) = \frac{1}{8};\ \ P(0'/0) = \frac{15}{16};\ P(1'/0) = \frac{1}{16}.$$

Требуется определить априорную энтропию $H(X)$, безусловную энтропию $H(X')$ с учётом корреляции, энтропию двух символов $H(X,X')$ и энтропию перехода $H(X'/X)$.

Согласно формул (4.8), (4.17), (4.13), (4.14) получим:

1. Априорная энтропия:

$$H(X) = \frac{3}{4}\log\frac{4}{3} + \frac{1}{4}\log 4 = \left(\frac{3}{4}+\frac{1}{4}\right)\log 4 - \frac{3}{4}\log 3 = 2 - \frac{3}{4}\cdot\frac{10\lg 3}{3} = 0.8,$$

2. Безусловная энтропия:

$$P(1') = P(1)\cdot P(1'/1) + P(0)\cdot P(1'/0) = \frac{3}{4}\cdot\frac{7}{8} + \frac{1}{4}\cdot\frac{1}{16} = \frac{43}{64},$$

$$P(0') = 1 - \frac{43}{64} = \frac{21}{64}, \qquad H(X') = \frac{43}{64}\log\frac{64}{43} + \frac{21}{64}\log\frac{64}{21} = 0.95,$$

3. Энтропия переходов:

$$H(X'/X) = P(1)\cdot\left[P(1'/1)\log\frac{1}{P(1'/1)} + P(0'/1)\log\frac{1}{P(0'/1)}\right] +$$

$$+P(0)\cdot\left[P(1'/0)\log\frac{1}{P(1'/0)} + P(0'/0)\log\frac{1}{P(0'/0)}\right] = \frac{3}{4}\cdot\left[\frac{7}{8}\log\frac{8}{7} + \frac{1}{8}\log 8\right] +$$

$$+\frac{1}{4}\cdot\left[\frac{15}{16}\log\frac{16}{15} + \frac{1}{16}\log 16\right] = 0.413,$$

4. Энтропия 2-х символов:

$$H(X,X') = H(X') + H(X'/X) = 0.95 + 0.413 = 1.363.$$

5. Средняя энтропия одного символа:

$$H(X)=\frac{1}{2}H(X',X)=0.682<H(X)=0.8$$

оказалась меньше априорной или, что то же самое, энтропии при отсутствии корреляции.

Действительно, при декорреляции символов:

$$P(1'/0)=P(1'/1)=P(1')=P(1),\ P(0'/1)=P(0'/0)=P(0')=P(0),$$

при этом:

$$H(X'/X)=P(1)\cdot\left[P(1')\log\frac{1}{P(1')}+P(0')\log\frac{1}{P(0')}\right]+$$

$$+P(0)\cdot\left[P(1')\log\frac{1}{P(1')}+P(0')\log\frac{1}{P(0')}\right]=$$

$$=P(1')\log\frac{1}{P(1')}+P(0')\log\frac{1}{P(0')}=H(X')=H(X),$$

$$H(X',X)=H(X')+H(X')=2H(X),\ H_1(X)=\frac{1}{2}H(X',X)=H(X)=0.8\ .$$

Следовательно, декорреляция символов позволила поднять энтропию одного символа от 0.682 до 0.8. Если же после этого выровнять вероятности $p(1)=p(0)=0,5$, то получим максимальную энтропию:

$$H_{\max}=\log 2=1\,бит/симв\ .$$

Различие между действительной и максимальной энтропией приводит к понятию избыточности источника.

Вместе с тем надо подчеркнуть, что при анализе характеристик каналов передачи информации статистическая мера наиболее проста и часто достаточна для описания производительности (трафика) каналов. Она обладает такими достоинствами, как универсальность (характеризует источники различной физической природы), объективность (что определяется экспериментальным характером расчёта энтропии). В связи с этим статистическая мера и поныне остается основной при оценке информационных свойств систем передачи сообщений.

Рассмотрим возможность применения теории информации при расчете количества информации перерабатываемой человеком оператором. Хотя эти положения и расходятся с мнением авторов, но они являются иллюстративными по отношению к приведенным положениям теории информации.

4.5. Методы расчета количества информации, перерабатываемой человеком-оператором

Применение теории информации в эргономике обусловлено влиянием на деятельность человека неопределенности (энтропии) процессов и объектов управления.

В качестве меры неопределенности физической системы x, принимающей состояния x_i, $i = 1, 2, ..., n$, используется понятие энтропии.

Энтропия физической системы x при значениях состояний x_i определяется выражением (формула Шеннона):

$$H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i \ , \tag{4.19}$$

где P_i – вероятность наступления i-го состояния системы.

Энтропия системы, как это следует из формулы (4.19), тем больше, чем больше общее количество различных состояний и чем меньше отличаются друг от друга вероятности этих состояний.

При равновероятном появлении различных состояний, когда $p_i = P = 1/\text{n}$, энтропия максимальна и определяется выражением:

$$H(X) = -\log_2 P_i = \log_2 n \ , \tag{4.20}$$

Неопределенность системы уменьшается при получении каких-либо сведений об этой системе. Чем больше объем полученных сведений и чем они более содержательные, тем большая информация имеется о системе.

Поэтому количество информации следует измерять уменьшением энтропии той системы, для уточнения состояний которой предназначены эти сведения.

Если в результате получения сведений состояние системы стало полностью определенным, то количество полученной информации численно равно априорной энтропии системы.

Энтропия и количество информации измеряются в двоичных единицах информации (битах). Бит – это энтропия простейшей системы, имеющей два равновероятных состояния.

Количество информации, переработанной человеком-оператором, может существенно отличаться от количества поступившей информации (энтропии источника сообщений). Это отличие обусловлено, во-первых, тем, что, как и для технического канала, часть информации может быть потеряна за счет воздействия помех. Во-вторых, количество информации может увеличиться за счет использования дополнительной информации.

Общее количество информации, перерабатываемой оператором, определяется выражением

$$I_ч = H(X) + H_{доп} - H_{пом}, \tag{4.21}$$

где $H(X)$ – энтропия источника сообщений, или количество информации, получаемой оператором от информационной модели;

$H_{доп}$ – дополнительное количество информации, используемой оператором при решении задачи;

$H_{пом}$ – энтропия источника помех, или количество потерянной информации.

Энтропия источника сообщений может быть двух видов:

– энтропия выбора нужного сигнала;

– энтропия измерительного прибора.

Если оператору нужно выбрать один сигнал (или одно состояние сложного сигнала) из n возможных, то количество полученной при этом информации можно определить по формуле (4.20) при неравновероятном или по формуле (4.21) при равновероятном появлении сигналов.

Эти формулы оценивают энтропию взаимонезависимых сообщений.

Это наиболее простые ситуации в работе оператора. На практике возможны ситуации, когда предъявляемая последовательность сообщений (сигналов) обладает логической избыточностью. Это значит, что появление определенного сообщения (сигнала) может изменить вероятность появления следующего. Наличие логической избыточности равносильно уменьшению энтропии, так как появление определенного сигнала уменьшает неопределенность очередного состояния информационной системы.

На основе энтропийного анализа можно оценить также сложность работы оператора со стрелочными измерительными приборами, расположенными на информационной панели (считывание показаний, установка заданных значений). Энтропию сообщения от прибора определяют как

$$H = \log_2 \frac{x_{\max} - x_{\min}}{2\sigma}, \tag{4.22}$$

где $x_{\max}$ и $x_{\min}$ – максимальное и минимальное значение шкалы прибора;

σ – абсолютная погрешность считывания показаний с прибора, равная (обычно) половине цены деления шкалы прибора.

Различают следующие виды дополнительной информации, используемой оператором при логической обработке данных:

– информация, подлежащая запоминанию;

– информация, используемая при вычислениях;

– информация, используемая при проверке логических условий.

Если оператор должен запомнить в порядке поступления m символов (знаков, сигналов) из общего количества символов – K, то количество запоминаний информации определяется выражением:

$$H_{зап} = \sum_{i=1}^{v} \log_2 N_i + \log_2 R\,, \qquad (4.23)$$

Если же оператору не нужно запоминать порядок поступления сигналов, а необходимо помнить только сами сигналы, то количество информации оценивается по формуле:

$$H_{зап} = \log_2 C_k^m = \log_2 \frac{K!}{m!(K-m)!}\,, \qquad (4.24)$$

Формулы (4.22) – (4.24) справедливы только при равновероятном поступлении сообщений (сигналов).

Количество информации, используемое при простейших вычислительных операциях (сложении, умножении, делении), определяют по формуле

$$H_{выч} = -\sum_{i=1}^{v} \log_2 N_i + \log_2 R\,, \qquad (4.25)$$

где v – количество чисел, используемых для получения результата R;

N_i – максимально возможные значения используемых при вычислении чисел;

R – максимально возможное значение результата вычисления.

При проверке логических условий количество информации определяют по формуле

$$H_{лог} = \sum_{i=1}^{s} \log_2 n_i \qquad (4.26)$$

где s – количество проверяемых логических условий;

n_i – количество возможных альтернатив (исходов), возникающих при проверке i-го условия.

При выполнении управляющих действий (движений) оператором также используется дополнительная информация, определяемая энтропией двух видов:

– энтропия выбора нужного органа управления (выбора требуемого положения органа управления);

– энтропия движения руки к органу управления.

Энтропию первого вида (количество информации) определяют по формулам (4.19) или (4.20).

Энтропию второго вида (количество информации) определяют как

$$H_{дв} = \sum_{j=1}^{u} \log_2 \frac{2A_j}{W_j}, \qquad (4.27)$$

где A_j – амплитуда движения руки к j-му органу управления, т.е. расстояние, на которое перемещается рука;

W_j – размер j-го органа управления.

Приведенные выше формулы получены на основании следующих двух правил.

Первое правило учитывает, что величина энтропии оценивает сложность выбора одного состояния из нескольких возможных. Поэтому при нахождении энтропии в любом случае сначала нужно определить общее количество всевозможных состояний (вариантов выбора), а затем применить формулу (4.19) при равновероятных состояниях или формулу (4.20), если все варианты равновероятны.

Второе правило, используемое при определении энтропии, заключается в свойстве ее аддитивности. Это значит, что энтропия сложной системы равна сумме энтропий отдельных подсистем. Например, применение формулы (4.25) для подсчета количества информации основано на том, что при производстве вычислений человек v раз производит выбор одного числа из N_i возможных, а при получении результата – выбор одного числа из R возможных.

Третья составляющая, входящая в выражение (4.21) – энтропия потерь, обычно не определяется, поскольку для обученного оператора, делающего минимальное количество ошибок, она очень мала и практически не влияет на значение $I_ч$. Поэтому в практических расчетах $H_{пом}$ принимают равным 0.

Количество перерабатываемой человеком информации необходимо знать, чтобы решить три основные задачи:

1. Количество перерабатываемой информации – мера сложности решаемой задачи, поэтому таким способом можно сравнивать между собой различные виды операторской деятельности.

2. Знание количества информации позволяет решать задачу согласования скорости поступления информации с психофизиологическими возможностями оператора по ее приему и переработке, т.е. с его пропускной способностью.

3. Зная количество информации, можно определить время, которое затрачивает оператор на ее переработку, т.е. нормировать операторскую деятельность.

Время, необходимое оператору для решения задачи, определяют на основе методов теории информации. Применение теории информации основано на том, что это время прямо пропорционально количеству перерабатываемой информации. Следует иметь в виду, что различные виды информации перерабатываются оператором с различной скоростью. Поэтому формулу (4.19) можно записать в виде

$$\tau_{on} = a + \sum_{i=1}^{k} H_i / v_i \qquad (4.28)$$

где a – латентный период (скрытое время реакции человека на сигнал); H_i – количество информации i-го вида, перерабатываемой оператором; v_i – скорость переработки информации i-го вида.

Средние значения скорости переработки различной информации приведены в табл. 4.1.

Таблица 4.1

Средние показатели скорости обработки информации оператором

Вид перерабатываемой информации i	Скорость переработки информации V_i, бит/с
Выбор одного сигнала из нескольких возможных	4
Считывание информации с приборов	2
Информация, необходимая для запоминания	12
Информация, используемая при вычислениях	6
Проверка логического условия	4
Выбор нужного органа управления	4
Информация при выполнение движений	7

В эргономике применение теории информации для решения перечисленных выше задач связано с некоторыми трудностями:

1. Величина энтропии в теории информации зависит от длины физического алфавита сигналов и вероятностей их появления. Эти же показатели используются для оценки количества перерабатываемой человеком информации. В действительности человек пользуется собственным, внутренним алфавитом, отличным от физического, а субъективные вероятности сигналов для человека не всегда совпадают с объективными. Однако принципы формирования внутреннего алфавита сигналов и вероятностей их появления для человека еще до конца не раскрыты.

2. Теория информации занимается лишь стационарными процессами, статистические характеристики которых не меняются с течением времени.

Характеристики же человека в виду его обучаемости, утомляемости, действия различных факторов беспрерывно меняются во времени.

1. Теория информации не учитывает смысловую сторону информации, ее ценность и значимость. На деятельность же оператора оказывают влияние не только статистические характеристики сигналов, но их смысл и значение для оператора.

2. Различные виды информации неравноценны для оператора. Поэтому при расчетах необходимо учитывать их «удельный вес», неодинаковое влияние на результаты деятельности оператора. Примером такого учета является формула (4.28), в которой сделана попытка учесть различия в скорости переработки различных видов информации.

Несмотря на эти трудности, информационные методы имеют большое практическое значение на ранних этапах проектирования систем «человек – техника – среда» (СЧТС). Зачастую только с их помощью можно количественно оценить параметры и показатели деятельности оператора.

Пример 4.8. На рис. 4.5 показана информационная панель, с которой работает оператор. На панели расположены 4 индикаторных лампочки Л1 – Л4, 4 измерительных прибора И1 – И4 с диапазоном измерения от 0 до 10 условных единиц, 2 трехпозиционных переключателя П1 и П2 и 2 десятипозиционных переключателя П3 и П4.

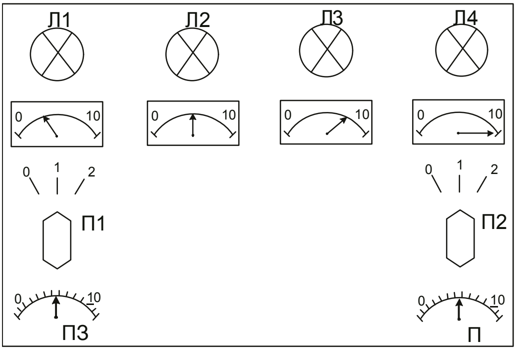


Рис. 4.5. Схема информационной панели

При поступлении входящего сигнала загорается одна из лампочек Л1 – Л4 (загорания лампочек равновероятны). Обнаружив загоревшуюся лампочку, оператор считывает показания соответствующего измерительного прибора (Л2 – И2). В зависимости от четного или нечетного номера загоревшейся лампочки оператор выбирает соответственно четную или нечетную группу переключателей, с помощью которой устанавливает новое значение измеряемой величины, в три раза большее отмеченной измерительным прибором.

Определить общее количество воспринимаемой и перерабатываемой оператором информации при появлении одного входящего сигнала и соответственно затрачиваемое при этом время.

Известно, что расстояние от оператора до информационной панели составляет 64 см, а расстояние между переключателями внутри группы равно 6 см. Скорости переработки информации оператором по видам информации приведены в табл. 4.1. Латентный период принимаем равным $a = 0,2$ $H_{пом}$ – энтропия источника помех (или количество потерянной информации) не учитывается.

Общее количество информации определяется по формуле (4.21). Необходимо рассчитать ее составляющие.

Поскольку загорания лампочек равновероятны, то первая составляющая $H(X)$ определяется по формулам (4.20) и (4.24).

$$H_1(x) = \log_2 4 = 2\ , \quad H_{пр}(x) = \log_2 \frac{x_{\max} - x_{\min}}{2\sigma} = \log_2 \frac{10-2}{2 \cdot 0,5} = 3,32 \text{ бит.}$$

Таким образом, $H(X)$ – энтропия источника информации или количество информации, получаемой оператором от информационной модели равно:

$$H(x) = 2 + 3,32 = 5,32 \text{ бит.}$$

Дополнительная информация $H_{доп}$ определяется таким образом.

$$H_{зап} = \log_2 C_4^1 = \log_2 \frac{4!}{1! \cdot 3!} = 2 \text{ бит.} \qquad H_{лог} = \log_2 2 = 1 \text{ бит.}$$

$$H_{дв} = \log_2 \frac{2A}{W} + \log_2 \frac{2A'}{W} = \log_2 \frac{2 \cdot 64}{4} + \log_2 \frac{2 \cdot 6}{4} = 6,58 \text{ бит.}$$

$$H_{выч} = \log_2 10 + \log_2 3 + \log_2 30 = 9,81 \text{ бит.}$$

$$H_{о.у} = \log_2 3 + \log_2 10 = 4,9 \text{ бит.}$$

Общая дополнительная информация равна:

$$H_{доп} = 2 + 1 + 6,58 + 4,9 = 24,29 \text{ бит.}$$

Общее количество воспринимаемой и перерабатываемой оператором информации равно:

$$I_ч = H(x) + H_{доп} = 4,32 + 24,29 = 29,61 \text{ бит.}$$

Общее время восприятия и переработки оператором информации

$$\tau_{опт} = a + \sum_{i=1}^{k} H_i / v_i = 0,2 + 2/4 + 3,32/2 + 2/12 + 1/4 +$$
$$+ 6,58/7 + 9,81/6 + 4,9/4 = 6,33\ c.$$

ГЛАВА 5. ОСНОВЫ СИСТЕМНОГО АНАЛИЗА

Изучение вопросов работы человеко-машинных систем очень и очень часто сталкивается с проблемой обработки информации оператором. При этом интересен даже не сам процесс обработки, а то, сколько этой самой информации обрабатывает оператор. Попытки использования классических методов оценки объема информации, как правило, к успеху не приводят. Оцененные с их помощью размеры информации оказываются зачастую просто огромными, такими, что оператор теоретически не может с ними справляться. На самом деле, в процессе работы он отлично справляется с существующими потоками информации, и это не вызывает в ходе работы особых трудностей. В чем же дело? В ходе экспериментов стало ясно, что классические методы подсчета количества информации зачастую не работают. Они не работают особенно тогда, когда речь идет о количестве обработанной оператором информации, необходимом количестве информации для принятия решений оператором, подсчете информационной емкости устройств отображения информации, подсчете информационной насыщенности информационных моделей и ряде других задач, возникающих при эргономическом проектировании систем информационного обеспечения деятельности операторов.

Данная глава представляет собой реферативный материал на тему неклассических методов подсчета количества информации, который может оказаться полезным при решении описанных выше задач. Для полноты описания здесь же в качестве справочного материала приведены и основные положения классической теории информации. Однако основной задачей данной главы видится создание почвы для поиска новых и пересмотра существующих подходов к подсчету количества информации в процессе деятельности оператора, в ходе обработки информации и принятия решений, оценки информационных возможностей систем информационного обеспечения деятельности операторов, при эргономическом проектировании систем информационного обеспечения.

5.1. Системный анализ проблемы

Выполнение любого задания, исследования или проекта начинается с постановки задачи и чем более четко, более однозначно (более формализовано) эта задача будет поставлена, тем более качественно она будет решена.

Поэтому, прежде чем сформулировать задачу и сформулировать так, чтобы результаты ее решения соответствовали поставленной цели, необходимо проделать большую работу, связанную с предварительным анализом проблемы и всего того, что с ней связано. Что же такое “проблема” и как необходимо ее решать?

Человек в процессе своей практической деятельности постоянно взаимодействует с внешней средой. Это взаимодействие носит пассивный и активный характер и выражается так:

– в познании среды;

– в адаптации к среде;

– в воздействии на среду;

– в управлении средой.

Состояние системы и окружающей ее среды на какой-то момент или отрезок времени называется ситуацией. В качестве модели ситуации можно рассматривать определенное сочетание свойств системы и среды, которые можно охарактеризовать совокупностью контролируемых переменных (показателей). В том случае, когда значения этих показателей (выраженных в некоторых шкалах, желательно в сильных) находятся в допустимых по какому-то критерию пределах, ситуация оценивается как благоприятная. В противном случае можно говорить о неблагоприятной ситуации. Такая ситуация часто называется проблемной.

Проблемная ситуация – это такое состояние человека (системы) и среды, при котором неудовлетворенность существующим положением дел осознана определенным лицом, но не ясно, что следует предпринять для ее изменения. Такая ситуация порождает проблему.

Неудовлетворенность ситуацией носит двойственный характер.

С одной стороны, это отрицательное отношение к тому, что имеет место, но, по мнению данного лица, не имеет право на дальнейшее существование. Проблемы подобного типа иногда называют негативными.

Проблемы другого типа возникают, когда неудовлетворенность ситуацией основывается на стремлении человека получить нечто желаемое, то, что еще не существует, но, по его мнению, должно быть. Эти проблемы можно назвать позитивными. К таким проблемам относятся проблемы построения и совершенствование орудий труда, систем информационного обеспечения и другие. Эти проблемы возникают, когда стремятся подогнать реальность под некоторую прагматическую (нормативную) модель, которая адекватно отражает желаемую ситуацию. Системный подход к понятию «проблема» отражен в приведенных ниже различных формулировках, отличающихся различным уровнем формализации и дающих возможность взглянуть на проблему с различных точек зрения.

1. Проблема (гр. problema – задача) - это сложный теоретический или практический вопрос, требующий изучения и решения.

2. Проблема – это осознание субъектом невозможности разрешить трудности и противоречия, возникающие в данной ситуации, средствами

наличного знания и опыта. Проблема осознается как такая противоречивая ситуация, в которой имеют место противоположные позиции при объяснении одних и тех же объектов, явлений и процессов или отношений между ними.

3. Проблема – это неблагополучное положение в какой-либо области человеческой деятельности, т.е. расхождение между требуемым (ожидаемым, желаемым) и фактическим состоянием системы или результатов ее функционирования.

4. Проблема – это осознание и формулировка неудовлетворенностей в отношениях системы и среды. Эта неудовлетворенность может проявляться в трех формах:

– неудовлетворенность воздействия внешней среды на систему (неудовлетворенность по входу системы);

– неудовлетворенность воздействия системы на внешнюю среду (неудовлетворенность по выходу системы);

– неудовлетворенность внутренним состоянием системы (неудовлетворенность по элементному составу, структуре, функциям, процессам и т.п.).

Этап осознания проблемы как некоторой иерархии неудовлетворенностей должен заканчиваться формулировкой проблемы, т.е. вербальной моделью проблемы.

Прежде чем формулировать проблему, необходимо четко представить для какого лица, организации или системы эта модель проблемы будет разрабатываться и где границы этой системы, чтобы были ясны те отношения (связи) с внешней средой, в которых какие-то процессы протекают не так как надо. При формулировке проблемы полезно предварительно получить ответы на следующие системные вопросы.

1. Кто (конкретное лицо, организация или другой системный объект, например, информационная система) неудовлетворен существующей ситуацией?

2. Что именно не удовлетворяет, какова иерархическая структура и ранжированная значимость этих неудовлетворенностей?

3. В какой среде осознается эта неудовлетворенность? Что представляет собой эта среда (каков ее состав и структура, какие процессы в ней протекают и какие релевантные факторы определяют ситуационное состояние среды)?

4. Как констатируемые неудовлетворенности соотносятся с общепринятой в данной культурной среде системой ценностей и с системой ценностей лица, формулирующего проблему?

5. Каковы пространственные и временные границы распространения этих неудовлетворенностей?

6. Выявлены ли эти неудовлетворенности в результате системного анализа или являются реакцией на какую-то ситуацию, т.е. каково соотношение объективного и субъективного при формулировке неудовлетворенностей?

7. Каковы возможные последствия сложившейся ситуации, если не принимать меры по ее нормализации?

На пути построения модели проблемы как системного объекта анализа существует много серьезных сложностей, некоторые из которых будут рассмотрены ниже.

Как уже было сказано, проблема зарождается в недрах некоторой системы, которую в системном анализе называют проблемосодержащей системой (ПС-системой), в отличие от системы, которая эту проблему будет решать и которую называют проблеморазрешающей системой (ПР-системой). Иногда эти системы выделяются на различных системных объектах-основаниях и разделены в пространстве и (или) времени; иногда объект-основание один, но системы различаются субстратным составом и структурой отношений.

Рассел Линкольн Акофф
(12.02.1919 – 29.10.2009)

В соответствии с системной методологией, любая проблемосодержащая система является подсистемой более высокого иерархического уровня, которая включает множество различных связанных между собой системных объектов (подсистем того же уровня).

Естественно, что все эти системы в той или иной мере взаимодействуют с рассматриваемой ПС-системой и число этих взаимодействий велико и разнообразно.

Поэтому, когда возникает неудовлетворенность между ПС-системой и средой, то она является следствием воздействия на систему большого числа взаимосвязанных факторов. Следовательно, то, что мы в определенной ситуации называем проблемой (выделенный набор неудовлетворенностей) на самом деле представляет только часть весьма сложного образования, которое Р. Акофф называет «клубком проблем» или проблематикой.

Проблематика – это совокупность взаимосвязанных проблем, имеющих место в проблемосодержащей системе и метасистеме, частью которой эта система является.

На практике не существует проблем в «чистом» виде: политических, экономических, военных, концессионных и т.п. Все они взаимосвязаны и интерпретация одних и тех же неудовлетворенностей как некоторой проблемы зависит от мировоззрения, интересов и квалификации того, кто эту проблему формулирует, т.е. создает вербальную модель проблемы.

Заметим, что даже если имеется общее согласие по определению проблемы, у системных аналитиков могут быть существенные расхождения во мнениях относительно масштабов и актуальности проблемы или степени ее взаимосвязи с другими проблемами (место проблемы в модели иерархии проблем и ранг проблемы на своем уровне иерархии).

Учитывая сказанное, системный аналитик подходит к исследованию проблемы с несколькими допущениями.

1. Проблемы взаимосвязаны и к их решению следует подходить целостно, иногда употребляют термин холистически (холизм (гр.) – философия целостности).

2. Проблемы не существуют сами по себе, а являются прямым результатом выбора, когда отдельные лица вырабатывают суждения об изменениях в состоянии дел. Другими словами, проблемы, по существу, носят еще и субъективный характер, так как они не существуют отдельно от субъективных интерпретаций тех, кто их определяет.

Пол Уильям Андерсон
(25.11.1926 – 31.07.2001)

3. Проблемы динамичны, во-первых, так как может быть столько решений, сколько формулировок проблем, а во-вторых, воздействует фактор времени и изменяется сама среда, в которой проблема зародилась, т.е. изменяется проблемосодержащая система.

Таким образом, основное предположение системного аналитика состоит в том, что проблема концептуально плохо структурирована. И это тем более верно, чем более тщательно и разносторонне изучается проблемная ситуация.

Очень хорошо эту мысль выразил писатель-фантаст П. Андерсон, сказав, что «проблема, сколь бы сложной она ни была, станет еще сложнее, если на нее правильно посмотреть».

Слабоструктурированность проблемы (особенно в ее первоначальной формулировке) объясняется тем, что к формулировке привлекаются лица, альтернативы, предпочтения и точки зрения которых могут быть множественными и конфликтующими, а вероятность того, что заданная альтернатива приведет к определенному результату или неизвестна, или не определена.

Основную задачу, которая решается при формулировании проблемы можно выразить так: не только избежать неправильных решений, но научиться задавать правильные вопросы, чтобы решать именно ту проблему, которую следует.

К сожалению, нельзя получить однозначный ответ на вопрос: «Как задавать правильные вопросы?» Однако новые технологии анализа предлагают концепции, которые системный аналитик может использовать, чтобы избежать получения точных ответов на неправильно поставленный вопрос, а получить хотя бы приблизительный ответ на правильно сформулированную проблему.

Одним и важнейших свойств проблемы является ее сложность.

Различают структурную и неструктурную сложности.

Структурная сложность определяется большим числом свойств или переменных (например, структура такой системы, как институт, завод, микрорайон города и т. п.).

Неструктурная сложность определяется качеством отношения между объектом и субъектом исследования (например, экономическое и социально-психологическое состояние общества).

Основное различие этих сложностей в том, что в первом случае приходится иметь дело хотя и с большим числом свойств и параметров, но которые выражаются в сильных квалиметрических шкалах (их можно измерить количественно), а во втором случае – эти свойства слабоструктурированы и либо вообще пока не измеримы, либо измеримы в слабых шкалах (шкалы наименований или порядка).

По второму классификатору различают объективную и субъективную сложности системных объектов.

Объективная сложность связана с сущностными свойствами анализируемого системного объекта, а субъективная сложность определяется особенностями субъекта анализа проблемы.

Очевидно, что при анализе такого нематериального слабоструктурированного объекта как проблема влияние субъективной сложности на построение модели проблемы является доминирующим, в отличии от построения системной модели такого, например, объекта как «спутниковая навигационная система». Поясним это более подробно.

Сложность спутниковой навигационной системы определятся большим числом (десятки тысяч) подсистем, агрегатов и элементов, объединенным единым замыслом, назначением и целью функционирования.

Сложность анализа и управления такими системами определяется огромным объемом постоянно обновляющейся информации. Однако эта информация имеет достаточно высокий уровень объективности, основанной на научных теориях, законах и высококачественной измерительной технике, а не на ограниченных квалиметрических способностях человека. Поэтому основной причиной, не позволяющей понижать сложность моделей управления такими системами являются ресурсные ограничения, например, финансовые, материальные или временные. Естественно, что для таких систем компонента объективной сложности существенно больше субъективной сложности, вносимой человеком как субъектом анализа системы. Наглядным примером объективной сложности являются сложности, возникающие при создании современного сверхзвукового самолета. Объективность сложностей подтверждает тот факт, что способы преодоления этих сложностей практически не зависят от страны-изготовителя и внешний облик новых истребителей напоминает близнецов.

Совершенно другая картина наблюдается, когда анализируется такой системный объект как проблема, которая по своей природе является абстрактным слабоструктурированным объектом без четких границ, состава и структуры, а свойства этого объекта (например, напряженность и значимость проблемы), выражаются, как правило, в слабых квалиметричских шкалах (шкала наименований или ранговая). В подобных системных объектах доминирует компонента субъективной сложности.

Указанные выше особенности проблемы как системного объекта анализа требуют более тщательного подхода к процедуре построения модели проблемы, которая должна носить итерационный характер.

Первоначальной формулировкой проблемы является формулировка заказчика решения проблемы, либо инициативного лица, которое взяло на себя ответственность впервые заняться этим вопросом.

Эта формулировка проблемы основывается, как правило, на некоторой интуитивной базе, не подкрепленной системным анализом вопроса и обладает следующими основными недостатками:

– нечеткость (неопределенность) формулировок;

– низкая степень формализации;

– ведомственность точки зрения на проблему (узость видения проблемы);

– излишней категоричностью;

– слабым представлением о сравнительной значимости проблемы на фоне других проблем в пространстве проблематики.

Исходя из этого, первоначальная формулировка проблемы, являясь вербальной моделью проблемы заказчика, должна быть использована системным аналитиком как исходная модель для построения более совершенной модели, обладающей большей мерой адекватности.

Поэтому следующим шагом построения модели проблемы является построения модели проблематики. Для этого необходимо исследовать по возможности более подробно ту метасистему (систему более высокого уровня), частью которой, по мнению аналитика, является рассматриваемая система. Чем более тщательно будет изучена эта метасистема, выявлены ее состав, структура, свойства и функции элементов, определены контролируемые и неконтролируемые переменные, тем более четко можно для изучаемой ПС-системы структурировать те неудовлетворенности, совокупность которых и будет являться базой построения модели проблемы.

После анализа проблематики и уточнения места и роли рассматриваемой проблемы в ее первоначальной формулировке, производят коррекцию этой вербальной модели проблемы.

Новая формулировка проблемы должна в доступной форме попытаться ответить на следующий комплексный вопрос: какие факторы, под воздействием каких сил и обстоятельств, управляемые какими людьми или организациями, преследующими какие цели приводят к ситуации, которую определенные субъекты деятельности воспринимают (классифицируют) как неудовлетворенность определенной степени, т.е. как проблему.

Полученная таким образом новая версия вербальной модели проблемы в обязательном порядке согласовывается с заказчиком или лицом принимающим решение (ЛПР) на ликвидацию или локализацию проблемы.

При формулировке проблемы полезно помнить высказывание Р.Акоффа: «В каждой проблемной ситуации существует некоторая совокупность относящихся к делу обстоятельств. Некоторые из них обычно кажутся очевидными. Чем более очевидными кажутся нам эти факторы, тем более тщательно необходимо проверять их истинность. Более вероятно ошибиться, принимая без доказательств кажущееся очевидным, чем требуя доказательств пусть даже сомнительных».

Из вышесказанного видно, что процесс формирования модели проблемы представляет собой сложный процесс, требующий проницательности, творческих способностей и взаимодействия участников анализа с целью выявления их ценностной ориентации, предпочтений и интерпретации любых имеющихся данных, относящихся к рассматриваемой проблеме. Такие процессы трудно поддаются формализации, однако современные технологии системного анализа позволяют в определенной мере формализовать некоторые этапы процесса анализа проблемы и повысить степень адекватности разрабатываемой модели проблемы.

5.2. Формализованная модель проблемы

Для более углубленного понимания процесса идентификации проблемы и построения ее формализованной модели, напомним и несколько расширим два важных системных понятия: состояние и ситуация.

Качество любого системного объекта может быть охарактеризовано совокупностью свойств.

Каждому из свойств может быть поставлена в соответствие некоторая мера, называемая показателем свойства, которая может квалиметрироваться в одной из сильных или слабых шкал.

Состоянием объекта будем называть совокупность значений показателей свойств качества объекта на фиксированный момент времени.

Создавая на системном объекте (как на модели-основании) интересующую нас систему, мы тем самым разделяем все обозримое множество системных объектов на два подмножества. Одно подмножество объектов связано с системой, а другое - с внешней средой. И система, и внешняя среда характеризуется в данный момент определенными состояниями.

Объединенное состояние системы и внешней среды на данный момент времени называется ситуацией.

Математическая модель ситуации может быть представлена в следующем виде:

$$C(t)=\left\{\Pi_c(t)_i\right\}\cup\left\{\Pi_{cp}(t)_j\right\}, \tag{5.1}$$

где $\left\{\Pi_c(t)_i\right\}$ - подмножество показателей свойств качества системы на момент времени t;

$\left\{\Pi_{cp}(t)_j\right\}$ - подмножество показателей свойств качества внешней среды на момент t.

Для каждого из показателей свойств качества системы и внешней среды выделяется область значений показателя, которая по какому-то критерию не соответствует определенному нормативу, и называется критической областью. В свою очередь, область значений показателя, соответствующая требуемым или желаемым значения, будем называть благоприятной областью.

Разделение всей области существования значений показателя на критическую и благоприятную области осуществляется с помощью одной или двух критических точек.

Как уже было сказано выше, категория «проблема» вводится как обобщенная характеристика неудовлетворительной ситуации.

Для того, чтобы формализовать проблему, необходимо выделить:

– подмножество нежелательных свойств системы и внешней среды;

– подмножество свойств системы и внешней среды, значения показателей которых находятся в критических областях (критические свойства);

– подмножество тех свойств системы и внешней среды, которые на данный момент отсутствуют, но которыми бы, по мнению ЛПР, они должны обладать (с заданным уровнем значений показателя свойства), т.е. желаемые свойства.

Численность полученного множества свойств может быть уменьшена, если отранжировать рассматриваемые свойства системы и среды по какой-то заданной системе критериев и ввести ограничение (меньше 1) на сумму их коэффициентов значимости, например, не более 0,9, или 0,99.

Таким образом, в качестве формализованной модели проблемы рассматривается совокупное множество нежелательных свойств, располагаемых и желаемых свойств рассматриваемой системы, показатели одних из которых находятся в критических областях (критические свойства), а показатели других (желаемых) свойств должны находиться в благоприятных областях. Такое множество свойств ПС-системы будем в дальнейшем называть проблемно-ориентированными свойствами или проблемой для рассматриваемой системы.

В литературе по системному анализу иногда говорят о формальном и системном подходах к формулировке проблемы.

Формальный подход применим к проблемам, которые иногда в литературе называют «жесткими». Такие проблемы можно достаточно хорошо структурировать и формализовать.

В отличие от жестких существуют «рыхлые» проблемы. Это проблемы слабоструктурированные и плохо выражаемые на естественном языке. Как правило, большинство проблем, с которыми приходится иметь дело на практике, относятся к категории рыхлых.

Формальный подход по своей сущности сводит «рыхлую» проблему к «жесткой». Однако опыт системного анализа показывает, что сводить «рыхлую» проблему к «жесткой» более опасно, чем наоборот. Если при переводе проблемы из «жесткой» в «рыхлую» исследователь частично отказывается от некоторой полезной информации, то в первом случае («рыхлая» в «жесткую») привносится ложная информация, которая вводит в заблуждение и может привести к неправильной оценке ситуации и неправильным решениям.

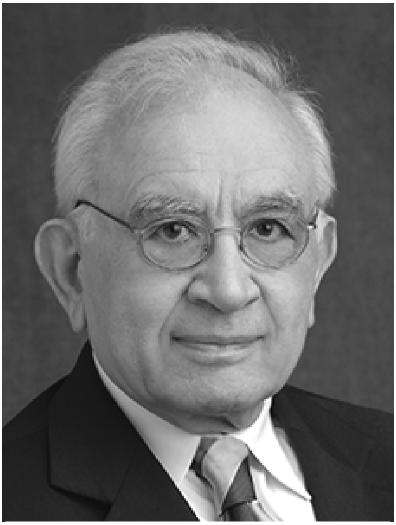
Опасность формального подхода заключается в том, что он часто сужает проблему до нереальной (нежизненной) схемы, создает видимость наукообразия глубины и точности системного анализа.

Системный подход к построению модели проблемы, в отличие от формального, обладает большей гибкостью и способствует повышению меры адекватности модели проблемы.

Основная цель системного подхода заключается в совершенствовании определения и структурирования проблемы с тем, чтобы научиться правильно задавать такие вопросы, ответы на которые позволят в той или иной мере решить проблему Неверно сформулированная проблема - это формулировка проблемы, способствующая подавлению симптомов, а не устранению причин, порождающих их.

5.3. Декомпозиция проблемы

Системный подход к исследованию такого сложного объекта как проблема позволил выделить ряд компонент (классов или групп) факторов, взаимодействие которых и образует проблемную ситуацию. Т. Саати предложил ввести в рассмотрение следующие группы факторов, которым для удобства идентификации присвоены условно следующие наименования (шкала наименований).



Томас Лори Саати
(род. 18.07.1926)

1. Факторы – группы факторов, характеризующих различные свойства отношений в государстве : политические, экономические, социальные, этнические, экологические и др., взаимодействие которых и образует проблему для данной ПС-системы.

2. Силы – факторы, с помощью которых меняется состояние вышеупомянутых факторов: наука, техника, технологии, культура, религия и т.п.

3. Актеры – люди, организации, партии, движения и т.п., которые мАнипулируют силами для воздействия на факторы.

4. Цели актеров, которые мотивируют действия актеров, влияющих с помощью сил на факторы.

5. Политики – способы поведения актеров в процессе достижения целей или способы использования ими средств достижения цели.

Для того, чтобы разобраться в структуре взаимодействия вышеперечисленных факторов необходимо вначале их структурировать и декомпозировать, а затем ранжировать по степени влияния друг на друга и на состояние ПС-системы.

Проанализировав характер взаимодействия факторов, необходимо задать себе следующие вопросы:

– какие последствия для ПС-системы будут иметь место, если степень взаимодействия факторов будет соответствовать высказанным экспертами приоритетам, т.е. в каких возможных состояниях может оказаться ПС-система, если в течении рассматриваемого времени характер взаимодействия факторов будет сохраняться?

– какое состояние ПС-системы считать желательным?

– какой способ решения проблемы выбрать?

– какую проблеморазрешающую систему для этого выбрать?

– какая стратегия ПР-системы должна быть реализована?

5.4. Сценарии

Для того, чтобы оценивать варианты (альтернативы) возможных, желаемых и требуемых состояний ПС-системы вводят в рассмотрение понятие "сценарий".

В системологической литературе встречается различные определения этого понятия, приведем лишь некоторые из них.

Сценарий – это система предположений о течении изучаемого процесса, на основе которой разрабатывается один из возможных вариантов прогноза.

По Гуревичу под сценарием понимают логическое правдоподобное описание событий с установлением примерного времени их наступления и связей, в результате которых данные события могут произойти.

По Саати сценарий – это описание определенной идеи (в виде, например, какого-то проекта) или рассматриваемого объекта (например, энергосистемы города) вместе с адекватной оценкой взаимодействий с факторами окружающей среды, а также социальными, политическими, экономическими и технологическими и др.

Сценарий должен по возможности однозначно выявить взаимодействие и влияние этих факторов с целью достижения убедительного описания состояния описываемого объекта при различных возможных предположениях. При построении сценария следует избегать чрезмерных полетов воображения и не входить в область фантастических предположений.

Сценарии могут описывать систему в различных состояниях. Их можно классифицировать следующим образом:

– исходные сценарии;

– логически возможные или вероятные сценарии;

– желаемые сценарии;

– требуемые сценарии.

При написании сценария пытаются воссоздать одно из состояний системы через количественное или качественное состояние ее компонент и динамики их взаимодействия. Сценарий обычно содержит вербальное описание некоторых взаимодействий, которые характерны для рассматриваемой ситуации.

Наиболее тщательно должны выписываться сценарии логически возможные и желаемые. При написании этих сценариев пытаются установить, как, исходя из существующей или заданной обстановки, начнут разворачиваться события в будущем. Особое внимание уделяется критическим точкам, после которых события могут разворачиваться в том или ином направлении.

На базе тщательно разработанного сценария можно разделить все факторы, относящиеся к будущему, на основные и второстепенные и сформулировать главные цели решения проблемы.

Кроме состояния ПС-системы с помощью сценариев можно описывать стратегии ее рационального поведения или поведения ПР-системы при реализации намеченной операции. Сценарии таких стратегий представляются в виде проектов или программ, содержащих детальное описание способов и средств достижения нужного состояния системы, привязанных к определенным временным этапам.

Для каждого из перечисленных видов сценариев рассматриваются несколько альтернативных вариантов, которые называют контрастными сценариями. Контрастные сценарии должны отличаться друг от друга каким-то одним или несколькими признаками, позволяющими экспертам высказывать достаточно убедительные суждения о степени их различия при сравнении по какому-то критерию.

Различные контрастные сценарии удобно описывать в терминах приращения показателей одного из основных свойств рассматриваемой системы или внешней Среды. Основное ценное свойство контрастного сценария должно состоять в том, что в нем следует по возможности точно выразить утверждения, в которых содержатся предложения о достижимости. Каждый контрастный сценарий описывает диапазон предположений, которые в совокупности составляют выпуклую оболочку возможных будущих вариантов.

По совокупности контрастных сценариев можно агрегировать **о**бобщенный сценарий, который представляет собой некоторую средневзвешенную смесь контрастных сценариев.

Алгоритм построения обобщенного сценария может быть представлен в виде последовательности следующих процедур.

1. Выбирается некоторая совокупность свойств исследуемого объекта $\{q_i\}$, с помощью которых можно сформировать модель системы, т.е.

построить контрастные сценарии для одного из указанных выше сценариев (возможных или желаемых).

2. Считая исходное состояние меры каждого из рассматриваемых свойств за «нуль отсчета» (в момент T_0), оценивают экспертным путем степень изменения меры каждого из свойств для всех рассматриваемых контрастных сценариев на прогнозируемый момент времени $T_{пр}$.

Квалиметрирование изменений показателей свойств (калибровка) производится по психометрической шкале Саати (табл. 5.1).

Таблица 5.1

Психометрическая шкала

Степень и направление изменения показателя свойства	Знак и мера измененияпоказателя
Значение не изменяется	0
Небольшое (слабое) увеличение (уменьшение)	+2 (-2)
Большое (сильное) увеличение (уменьшение)	+4 (-4)
Очень большое (значительное) увеличение (уменьшение)	+6 (-6)
Абсолютное (максимально возможное) увеличение (уменьшение)	+8 (-8)
Промежуточные значения	+1 (-1), +3 (-3), +5 (-5), +7 (-7)

Результаты экспертного оценивания тенденций изменения показателей свойств объекта заносят в табл. 5.2.

Таблица 5.2

Изменения показателей свойств объекта

	СЦ-1 (s_1)	СЦ-2 (s_2)	...	СЦ-j (s_j)	...	СЦ-k (s_k)	
n_1	$g_{11} = +3$	$g_{12} = -2$	...	g_{1j}	...	g_{1k}	$П_1^*$
n_2	$g_{21} = -4$	$g_{22} = +8$	...	g_{2j}	...	g_{2k}	$П_2^*$
...	...	...	...	...	...	...	...
n_i	$g_{i1} = +5$	$g_{i2} = -3$	...	g_{ij}	...	g_{ik}	$П_i^*$
...	...	...	...	...	...	...	...
n_m	$g_{m1} = -1$	$g_{m2} = +7$	...	g_{mj}	...	g_{mk}	$П_m^*$

По значениям табл. 5.2 рассчитывают средневзвешенные значения тенденций изменения показателей свойств объекта $П_i^*$ следующим образом:

$$П_i^* = \sum_{j=1}^{k} s_j \cdot g_{ij} \,, \tag{5.2}$$

где s_j – коэффициент значимости (предпочтительности) j-го сценария, полученный по методу анализа иерархий;

g_{ij} – мера степени изменения i-го свойства по j-му сценарию.

По совокупности средневзвешенных значений показателей свойств объекта $\{ П_i^* \}$ формулируют вербальную модель обобщенного сценария, который может реализоваться в результате взаимодействия рассмотренных групп факторов с учетом тех приоритетов (коэффициентов предпочтительности), которые получены по оценкам экспертов.

Для того чтобы в дальнейшем можно было каким-то образом сравнивать различные обобщенные сценарии, вводят меру сценария, которая представляет собой средневзвешенную меру обобщенных переменных, описывающих рассматриваемую ПС-систему:

$$M_{oc} = \sum_{i=1}^{m} b_i \cdot П_i^* \,, \tag{5.3}$$

где b_i – коэффициент значимости (компонента вектора приоритетов) i-ой переменной из табл. 5.2, полученная по экспертным оценкам с помощью матрицы парных сравнений в шкале Саати; $П_i^*$ - переменная из табл. 5.2.

Так как будущее определяется под воздействием различных сил и интересов (каждая старается реализовать свои цели), при синтезе различных сценариев в обобщенный должны быть приняты во внимание:

– актеры, влияющие на будущее;

– цели актеров и их отдельные линии поведения, которых они будут придерживаться для осуществления в каждом сценарии своих целей.

Таким образом, процесс построения обобщенного сценария должен отражать приоритеты актеров в соответствии с их важностью, осуществляя с достаточной полнотой части, составляющие каждый сценарий.

Процедура построения сценария включает такие основные этапы.

1. Определение проблемосодержащей системы, внешних и внутренних ограничений, а также идентификацию подсистем.

2. Построение иерархической структуры подсистем и идентификация регулируемых компонент (управляемых переменных).

3. Определение состояний системы, моделирование истории развития.

4. Моделирование развития сценария во времени (исторический анализ) для освещения эволюции системы и влияние этих процессов на характеристики, одновременно с исследованием внутренней динамики модели.

5. Определение целей сценария с обсуждением их значимостей.

6. Выбор типов сценариев.

7. Разработка основных данных прошлой, настоящей и будущей информации.

8. Идентификация структурных компонент и факторов, обеспечивающих равновесие и тенденции развития системы.

9. Исследование «трений» (препятствий), присущих рассматриваемым функциональны механизмам системы.

10. Анализ механизмов регулирования (точек управления) в системе и их связей.

11. Критический анализ и усовершенствование сценариев посредством проверки ограничений, неравновесности, трений, сил, противоречий, вмешательство механизмов регулирования, а также устранение противоречий, которые влияют на выживание системы.

12. Создание усовершенствованного сценария.

Таким образом, декомпозиционная модель проблемы включает следующие взаимосвязанные компоненты (рис 5.1):

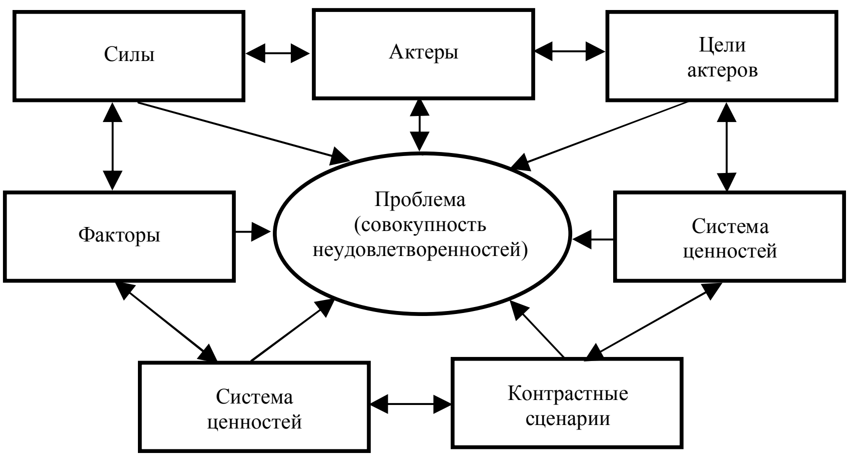


Рис 5.1. Декомпозиционная модель проблемы

1. Структурированные неудовлетворенности в виде совокупности нежелательных, критических и желаемых свойств проблемосодержащей системы (ПС-системы).

2. Группы факторов (факторы, силы, актеры, цели актеров), взаимодействие которых приводит к появлению неудовлетворенностей.

3. Контрастные сценарии, которые позволяют исследователю анализировать взаимодействие факторов и высказывать суждения о возможных тенденциях в динамике проблемы.

5.5. Решение проблемы

5.5.1. Концептуальные подходы к решению проблемы

Правильный выбор целей, ПР-системы и стратегий ее поведения зависит как от качественного анализа факторов, воздействие которых и формирует выявленные неудовлетворенности, т.е. проблему, так и от оценки значимости (важности) влияния каждого из рассматриваемых факторов на проблему.

В самом общем виде решить проблему - это значит перевести проблемосодержащую систему из проблемной ситуации в желаемую ситуацию, при которой ликвидируются нежелательные свойства, показатели критических свойств ПС-системы переводят в благоприятные области и появляются новые желаемые свойства.

Выше были сформулированы понятия негативных и позитивных проблем. С учетом этого могут быть различные подходы к решению проблем.

При решении негативных проблем основное усилие направляется на устранение того, что нежелательно в данной ситуации, т.е. решение проблемы ориентировано на результаты анализа прошлого опыта. Такая форма решения проблемы получила в системологической литературе название ретроспективной. С учетом введенного в рассмотрение понятия «критическое свойство» можно сказать, что ретроспективное решение это такое решение проблемы, при котором ПР-система обеспечивает перевод показателей критических свойств из критических областей в благоприятные. Недостатком такого подхода к решению проблемы является то, что акцент внимания на устранение недостатков прошлого снижает вероятность учета последствий предпринятых действий и тем самым увеличивает вероятность появления новых проблем.

При решении позитивных проблем усилия направлены на достижения желаемой цели, сформулированной в терминах приближения к какому-то идеалу, то есть к созданию у ПС-системы новых желаемых свойств. Такое решение проблемы, называемое перспективным, уменьшает вероятность просчета в возможных последствиях.

Ретроспективная форма решения проблемы направлена на поддержание функционирования системы на требуемом уровне, а перспективная- на развитие системы, т.е. на обеспечение достижения измененных целей системы.

Для того, чтобы можно было с какой-то степенью объективности принимать решение на выбор способа решения проблемы и оценивать результаты процесса решения необходимо ввести какую-либо меру степени решения проблемы.

Напомним, что формализованная модель проблемы представляет собой множество располагаемых нежелательных, критических и желаемых свойств.

Степень решения проблемы можно связать с долей ликвидированных нежелательных и критических свойств и долей появления желаемых свойств. Желательно также, чтобы учитывалось не только число тех или иных свойств, но и их значимость с точки зрения достижения ПС-системой своих целей.

Если проблема рассматривается как совокупность нежелательных, критических и желаемых свойств ПС-системы, то в качестве меры степени решения проблемы может быть использовано следующее выражение:

$$M_{pn} = v_p \cdot m_p + v_n \cdot m_n \,, \tag{5.4}$$

где v_p и v_n - коэффициенты, характеризующие долю реализации ретроспективной и перспективной компоненты решения проблемы, определяемые экспертами по шкале Саати и удовлетворяющие следующему условию: $v_p + v_n = 1$; m_p и m_n - меры, характеризующие соответственно степень решения ретроспективной и перспективной компоненты проблемы, которые определяются по выражениям:

$$m_n = a_н \sum_{i,z} E_i + a_{кр} \sum_{i,k} S_i \,, \tag{5.5}$$

$$m_n = \sum_{j,h} D_j \,, \tag{5.6}$$

где E_i, S_i, D_j - коэффициенты значимости соответственно нежелательных, критических и желаемых свойств ПС-системы, определяемые экспертным путем (по шкале Саати);

z, k, h - соответственно число нежелательных, критических и желаемых свойств в рассматриваемой модели проблемы;

$a_н$, $a_{кр}$ - соответственно коэффициенты ликвидности нежелательных и критических свойств ПС-системы, назначаемые или оцениваемые экспертным путем, причем, $a_н + a_{кр} = 1$.

В общем случае, при наличии реальных проблем приходится реализовывать некоторый комбинированный способ, включающий в определенной пропорции ретроспективную и перспективную компоненту решения. В свою очередь, ретроспективное решение проблемы обеспечивается определенным компромиссом между устранением долей нежелательных и критических свойств ПС-системы.

Оценка интегральной меры решения проблемы может меняться в зависимости от степени важности (с точки зрения целей ПС-системы) ретроспективной и перспективной компоненты для ЛПР. Соотношение степеней важности задается с помощью коэффициентов значимости этих компонент, определяемых экспертным путем по матрице парных сравнений с использованием шкалы Саати.

Р. Акофф рассматривает четыре способа решения проблемы:

– solution - решать проблему наилучшим для данной ситуации способом, т.е. способом, оптимальным по одному или нескольким критериям;

– resolution - с учетом возможных ресурсных ограничений, решить проблему частично, смягчив ее до допустимого предела;

– absolution - не решать проблему, надеясь на то, что она исчезнет сама собой;

– disolution - ликвидировать, растворить проблему, изменив условия или точку зрения, произведя в проблемосодержащей системе и (или) в ее окружении (или в модели) такие изменения, чтобы не только исчезла сама проблема, но и будущие (возможные) проблемы система смогла преодолеть самостоятельно (принцип самоорганизации).

5.5.2. Основные этапы подготовки решения проблемы

Выше было сказано, что на практике исследователь сталкивается не с какой-то отдельно взятой проблемой, а с множеством проблем. Поведение этого множества проблем в большей степени зависит оттого, как будут взаимодействовать между собой способы решения проблем, их составляющие, чем от того, как решается каждая проблема в отдельности.

Нужно помнить, что оптимальное функционирование частей не исключает возможности гибели целого. Поэтому, частные решения всей системы проблем (проблематики) лучше, чем полное решение каждой из проблем, взятых отдельно друг от друга.

Решение проблемы – это преобразование ПС-системы из исходного, неудовлетворительного состояния в желаемое или требуемое состояние.

Для оценки располагаемого, возможного или желаемого состояний анализируемой системы выделяются ее сущностные свойства и для каждого из них вводится некоторый квалиметрированный показатель. В зависимости от ситуации совокупность значений этих показателей в общем случае изменяется. Это изменение является результатом текущего взаимодействия среды и системы и могут иметь тенденцию как к росту, так и к уменьшению.

Для того, чтобы управлять процессом решения проблемы, необходимо оперировать с моделью ПС-системы, моделью ПР-системы и моделью ситуации.

Эффективность решения проблемы зависит от:

– правильного (адекватного) анализа проблемы, связанной с определенной ПС-системой;

– правильного выбора цели, достижение которой будет отождествляться с требуемым уровнем решения проблемы; в данном случае цель – это желаемое или требуемое состояние рассматриваемой ПС-системы;

– правильного выбора ПР-системы которая, обладая всеми необходимыми свойствами и ресурсами, способна перевести ПС-систему из располагаемого (проблемного) состояния в желаемое, т.е. достичь целей преобразования;

– правильного выбора способа (технологии) достижения цели, т.е. такого использования средств и ресурсов, которое минимизирует ресурсные затраты или максимизирует отношение эффект/затраты.

Правильный выбор целей, ПР-системы и стратегий ее поведения зависит как от качественного анализа факторов, воздействие которых и формирует выявленные неудовлетворенности, т.е. проблему, так и от оценки значимости (важности) влияния каждого из рассматриваемых факторов на проблему.

Для проведения анализа проблемы и выбора способа ее решения был разработан метод, получивший название метода анализа иерархий (МАИ). Т.Саати предложил агрегировать указанные на рис. 5.1 компоненты в некоторую иерархическую структуру, анализ которой с точки зрения приоритетов их взаимодействия позволяет выработать предложения по решению проблемы, т.е. наметить средства и способы достижения намеченных целей.

Как было сказано выше, для того, чтобы решить проблему необходимо правильно выбрать цель, отождествляемую с желаемым уровнем решения проблемы, и средства и пути достижения намеченной цели.

Анализ и выбор желаемых целей осуществляется в следующей последовательности:

– разрабатывается параметрическая модель ПС-системы;

– оценивается возможное логическое будущее состояние ПС-системы;

– оцениваются желаемые состояния ПС-системы и выбирается требуемое, т.е. выбирается цель операции для ПР-системы;

– оцениваются возможные стратегии достижения цели ПР-системой и выбирается требуемая.

Оценка логически возможного сценария состояния ПС-системы:

Вначале разрабатывается параметрическая модель проблемосодержащей системы (ПС-системы) в виде множества сущностных свойств системы, квалиметрированных в каких-либо шкалах (желательно в сильных).

Затем разрабатываются несколько контрастных сценариев, которые описывают возможные состояния ПС-системы в терминологии ее сущностных свойств.

Для того, чтобы оценить логически возможные взаимные приоритеты этих сценариев агрегируют по методу МАИ структуру иерархической модели, связывающую следующие компоненты (уровни) модели проблемы:

– фокус иерархии – единственный элемент самого верхнего уровня, который определяется (формулируется) как логическое будущее ПС-системы (нулевой уровень иерархии);

– факторы, которые с той или иной интенсивностью воздействуют на проблему, т.е. формируют проблему (1-ый уровень);

– силы, которые влияют на факторы и их динамику (2-ой уровень);

– актеры ПС-системы, которые с помощью сил влияют на факторы (3-ий уровень);

– цели актеров, которые определяют мотивацию активной деятельности актеров и через силы влияют на динамику факторов, взаимодействие которых и порождает некоторую логически обоснованную динамику проблемы (4-ый уровень);

– контрастные сценарии возможного логического будущего состояния ПС-системы (5-ый уровень);

– обобщенный сценарий логического будущего ПС-системы (6-й уровень).

Модель такой структуры представлена на рис. 5.2.

Далее формируют с помощью экспертов матрицы парных сравнений, по которым рассчитываются локальные приоритеты элементов каждого уровня по отношению к каждому элементу верхнего уровня.

Затем рассчитываются глобальнее приоритеты элементов самого последнего (нижнего) уровня, в роли которых и выступают контрастные сценарии.

На завершающем этапе исследования вопроса о логически возможном состоянии ПС-системы, который называется прямым процессом анализа, рассчитывают средневзвешенные показатели обобщенного сценария.

Анализ исходного (в рассматриваемый момент времени) состояния ПС-системы и ее логически возможного состояния позволяет исследователю сформировать для себя определенное представление о возможных способах решения проблемы и о той ПР-системе, которая способна эту проблему решить.

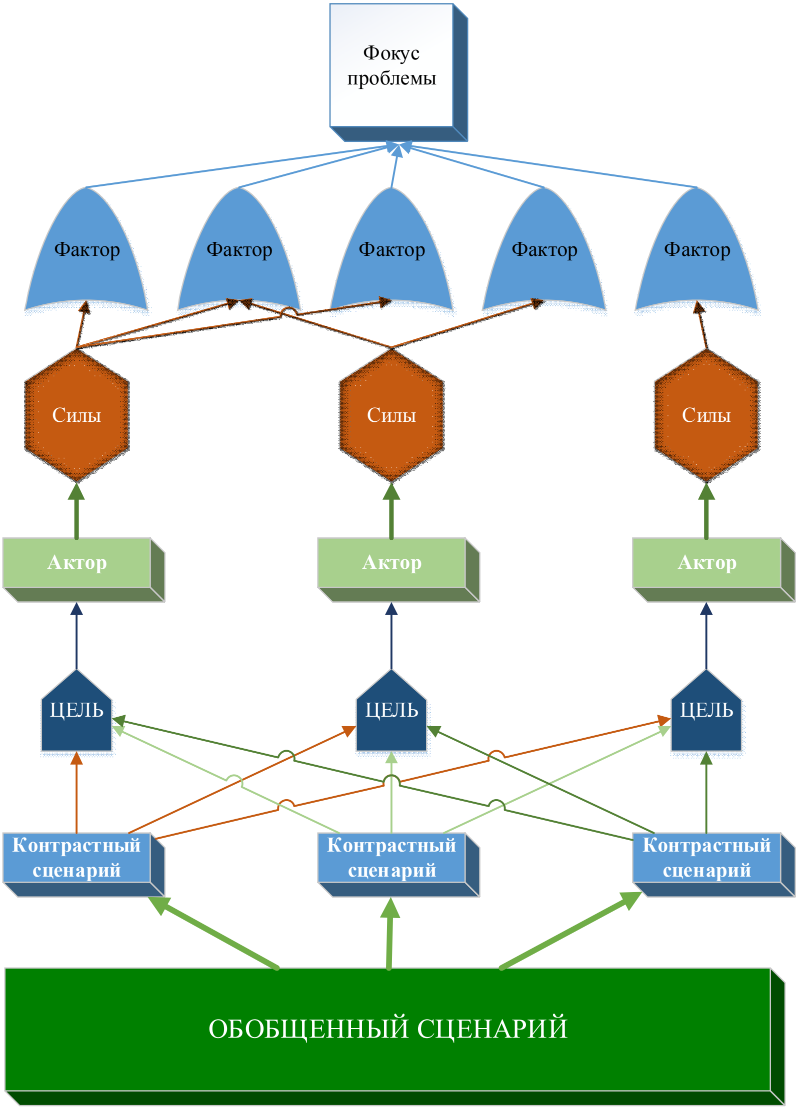


Рис. 5.2. Структура иерархии анализа проблемы

Способ решения проблемы будет определяться теми целями, которые ставит перед собой лицо, принимающее решение (ЛПР). Поэтому следующей важнейшей задачей, стоящей перед исследователем, является разработка модели целей для выбранной ПР-системы.

Выбор цели.

Построение модели цели реализуется в два этапа. На первом, наиболее ответственном этапе разрабатывается модель метацели (цели самого высшего иерархического уровня), которая представляет собой вербальную модель желаемого состояния ПС-системы, т.е. вербальную модель желаемого сценария.

На втором этапе разрабатывается декомпозиционная модель цели.

Для повышения степени объективности в принятии решения о выборе одного из альтернативных желаемых сценариев возможно применение различных методов, в частности можно применить метод анализа иерархий. Главная задача такого анализа – выявить и оценить приоритетные сценарии, реализация которых позволит решить поставленную задачу.

На рис. 5.2 представлена формальная иерархическая модель, которая позволяют оценить приоритеты рассматриваемых контрастных желаемых сценариев (проектов) по совокупности критериев, характеризующих выгоды (эффективность) и ущерб (затраты, потери) от реализации рассматриваемых сценариев.

После определения коэффициентов значимости для каждого проекта по первой и второй моделям определяют для каждого проекта отношение коэффициентов значимости i-го проекта по критерию «выгода» и по критерию «ущерб» и рассматриваемые проекты ранжируются по величине этих отношений. Приоритет отдается тому проекту, у которого величина этого отношения максимальная.

Выбор стратегии достижения цели.

Следующий этап исследования заключается в поиске рациональных способов (стратегий) достижения ПР-системой целей преобразования ПС-системы. Для этого агрегируется иерархическая модель, представленная на рис. 5.3, которая называется обратным процессом анализа. Фокусом этой модели является выбранный желаемый сценарий состояния ПС-системы, а элементами последующих (в порядке возрастания номера) уровней являются:

– неблагоприятные факторы, препятствующие реализации рассматриваемого сценария (нежелательные свойства ПС-системы или внешней среды);

– актеры ПР-системы, которые с той или иной степенью активности способны (или должны) воздействовать на негативные факторы, чтобы обеспечить реализацию желаемого сценария;

– цели актеров, мотивирующие их деятельность;

– альтернативные программы (стратегии), которые могут использоваться актерами для достижения запланированных целей.

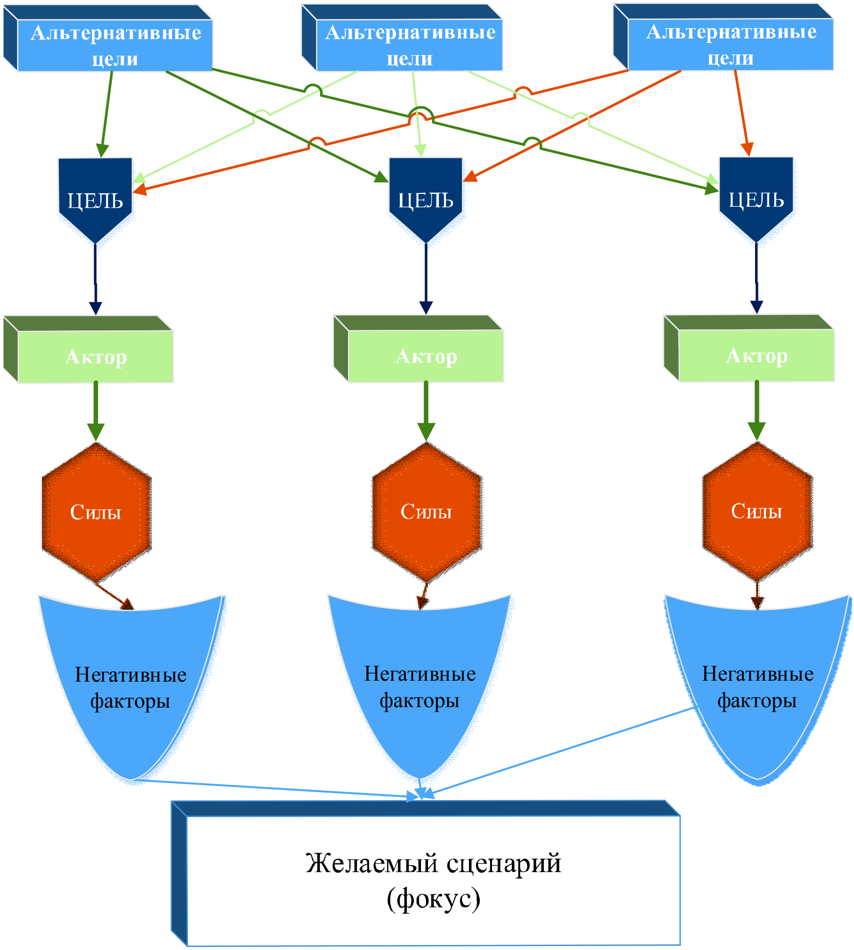


Рис. 5.3. Структура обратного анализа

После проведения экспертного оценивания элементов всех уровней иерархической модели и оценки локальных приоритетов рассчитывают вектор глобальных приоритетов альтернативных программ (стратегий).

Реализация прямого и обратного процессов анализа проблемы представляет собой первый замкнутый цикл анализа, представленный на рис. 5.4.

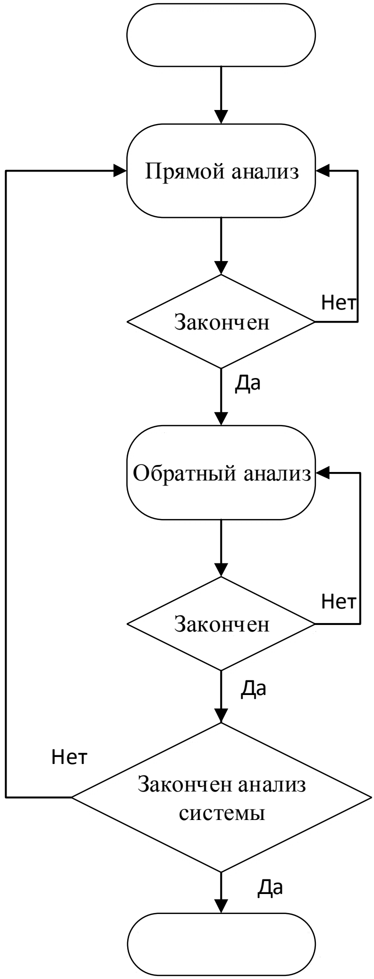


Рис. 5.4. Циклы анализа системы

Полученная на 1-й итерации анализа информация о приоритетности факторов, сил актеров и их целей вызывает необходимость в проведении 2-й итерации прямого процесса анализа, где уточняются или изменяются актеры, их цели и соответствующие приоритеты. Это уточнение приводит к изменению калибровки параметров модели ПС-системы и, как результат, к изменению параметров обобщенного сценария логического будущего. Исследователь сравнивает новый вариант логического будущего ПС-системы с желаемым будущим сценарием и, с учетом меры рассогласования, вносит соответствующие коррективы во вторую итерацию обратного процесса анализа, что, в свою очередь, приводит к изменению приоритетов программ (стратегий) достижения желаемого сценария. Итерационный процесс повторятся до тех пор, пока исследователь не добьется требуемой степени сходимости между возможным и желаемым логическим будущим ПС-системы.

5.5.3. Алгоритм анализа и решения проблемы

Блок-схема алгоритма решения проблемы показана на рис. 5.5. Основные пункты данного алгоритма описаны ниже.

1. Осознание неудовлетворенностей каким-либо субъектом анализа.
2. Диагностирование симптомов проблемы.
3. Идентификация проблемосодержащей системы (ПС-системы). Формирование подмножества контролируемых переменных.
4. Анализ проблематики.
5. Идентификация проблемы (нежелательные, критические и желаемые свойства ПС-системы).

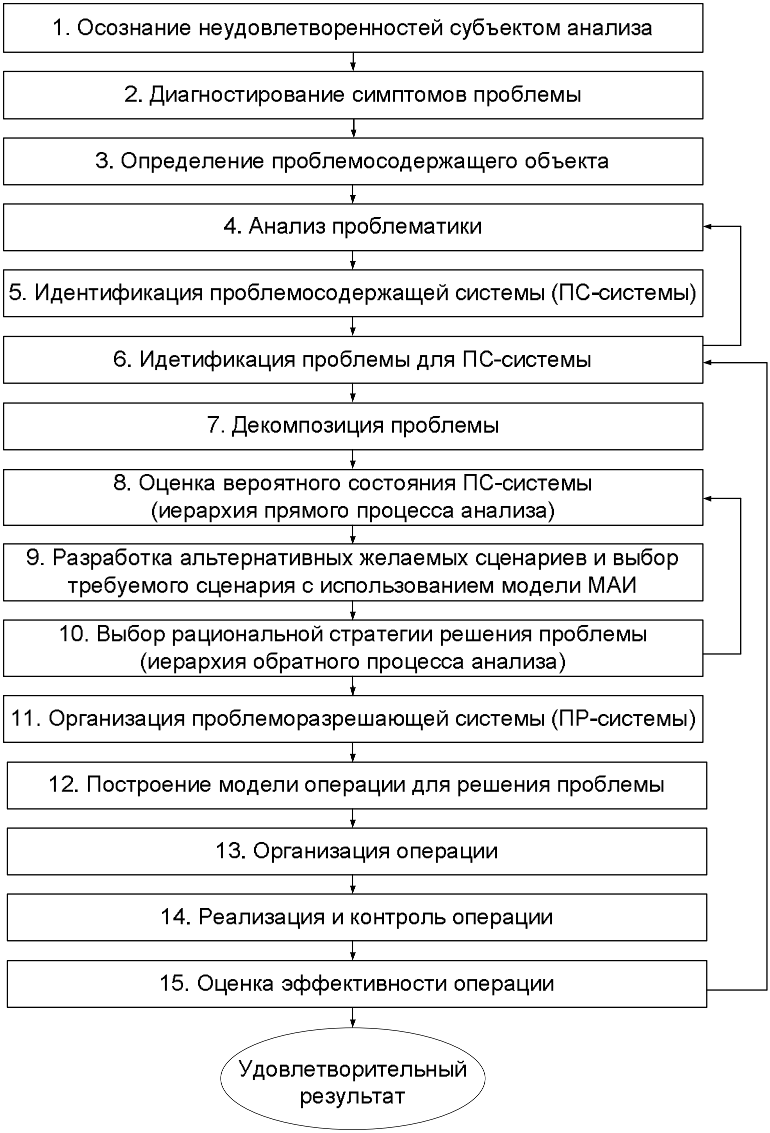


Рис. 5.5. Блок-схема алгоритма решения проблемы

6. Декомпозиция проблемы (факторы, силы, актеры, цели актеров, контрастные сценарии вероятного состояния ПС-системы).

7. Агрегирование и исследование иерархии прямого процесса анализа с целью построения обобщенного сценария вероятного состояния ПС-системы.

    7.1. Анализ факторов и сил, под воздействием которых зародилась и развивается проблема.

    7.2. Анализ степени влияния актеров, активно воздействующих на динамику проблемы.

    7.3. Оценка приоритетов целей актеров, под влиянием которых формируется мотивационная компонента их деятельности.

    7.4. Оценка приоритетов контрастных сценариев логически возможных состояний ПС-системы в среде рассматриваемой проблемы.

    7.5. Формирование обобщенного сценария возможного состояния ПС-системы.

8. Разработка контрастных желаемых сценариев.

9. Выбор требуемого сценария (требуемого состояния ПС-системы).

    9.1. Агрегирование иерархии модели ранжирования желаемых сценариев по критерию “выгода”.

    9.2. Агрегирование иерархии модели ранжирования желаемых сценариев по критерию “ущерб”.

    9.3. Ранжирование желаемых сценариев по критерию “выгода / ущерб”.

    9.4. Выбор требуемого сценария (стратегической цели решения проблемы).

10. Агрегирование и исследование иерархии обратного процесса анализа с целью нахождения рациональной программы или стратегии достижения требуемого сценария.

    10.1. Анализ степени влияния неблагоприятных факторов, препятствующих достижению выбранной цели.

    10.2. Оценка приоритетов актеров по степени их влияния на интенсивность воздействия факторов.

    10.3. Оценка коэффициентов значимости целей актеров ПР-системы.

    10.4. Расчет векторов локальных приоритетов рассматриваемых альтернативных программ (стратегий)достижения требуемого сценария.

    10.5. Выбор программы достижения цели.

11. Построение модели операции по достижению цели (планирование операции). Оценка условий проведение операции.

12. Организация операции, выбор (создание) ПР-системы.

13. Проведение операции и контроль этого процесса.

14. Оценка эффективности операции.

Глава 6. ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ

Этот раздел будет несколько необычно выглядеть в данной книге. Он будет самым коротким и самым спорным. Нам пришлось долго бороться с собой и убирать все лишнее в данном разделе. Лишней, как ни странно, оказалась математика. Точнее не математика, а формулы, описывающие различные методы и подходы.

Однако обойти данную область человеческого знания мы не могли. Поэтому посвятили данный раздел общим вопросам и подходам к решению проблемы выбора. Что послужило основанием для необходимости включения данного раздела?

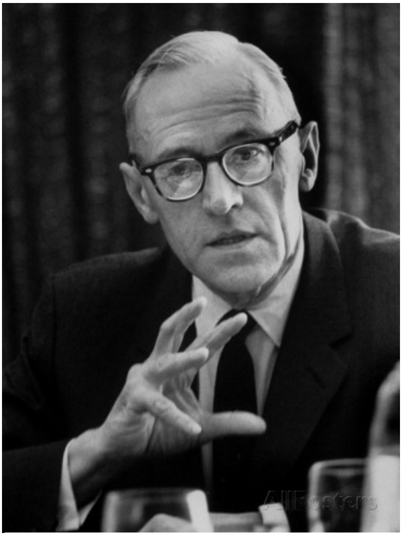
Выбор сам по себе представляет довольно сложную проблему. Ее решение в такой области как эргономика зачастую связано с опытом, риском и конкретными условиями. Однако практически все эти виды неопределённости уже исследованы в различных областях математики и применительно к различным задачам. Поэтому данный раздел снабжен отличным списком литературы основываясь на котором можно найти более полное описание математических методов необходимых в конкретных условиях. Тем более, что знание данных методов зачастую выходит за рамки профессиональной компетенции эргономистов. Поэтому им необходимо знать о возможности такого подхода и подходить к решению конкретных задач с привлечением математиков или инженеров.

Джон фон Нейманн
(28.12.1903 – 08.02.1957)

6.1. Основы процесса принятия решений

Искусство принятия наилучших решений, основанное на опыте и интуиции, является сущностью любой сферы человеческой деятельности. Наука о выборе приемлемого варианта решения сложилась сравнительно недавно, а математической теории принятия решений – около 50 лет назад.

Основы теории принятия решений разработаны Джоном фон Нейманом и Оскаром Моргенштерном. По мере усложнения задач появилось много различных направлений этой науки, которые имеют дело с одной и той же проблемой анализа возможных способов действия с целью нахождения оптимального в данных условиях решения проблемы.



Оскар Моргенштерн
(24.01.1902 – 26.07.1977)

Как самостоятельная дисциплина общая теория принятия решений (ТПР) сформировалась в начале 60-х годов XX века, тогда же была сформулирована основная цель этой теории – рационализировать процесс принятия решений. В последующие годы была создана и прикладная теория статистических решений, позволяющая анализировать и решать широкий класс управленческих задач, связанных с ограниченным риском – проблемы выбора, размещения, распределения и т.п.

В настоящее время теория принятия решений применяется преимущественно для анализа тех проблем, которые можно легко и однозначно формализовать, а результаты исследования адекватно интерпретировать. Так, например, методы ТПР используют в самых различных областях управления – при проектировании сложных технических и организационных систем, планировании, выборе программ развития экономики и энергетики регионов, организации новых экономических зон и т.п.

Необходимость использования подходов и методов ТПР в управлении очевидна: быстрое развитие и усложнение разнородных связей, выявление зависимости между отдельными сложными процессами и явлениями, которые раньше казались не связанными друг с другом, приводят к резкому возрастанию трудностей принятия обоснованных решений. Затраты на их осуществление непрерывно увеличиваются, последствия ошибок становятся все серьезнее, а обращение к профессиональному опыту и интуиции не всегда приводит к выбору наилучшей стратегии. Использование методов ТПР позволяет решить эту проблему, причем быстро и с достаточной степенью точности.

В задаче ТПР человек (или группа лиц) сталкивается с необходимостью выбора одного или нескольких альтернативных вариантов решений (действий, планов поведения). Необходимость такого выбора вызвана какой-либо проблемной ситуацией, в которой имеются два состояния: желаемое и действительное, а способов достижения желаемой цели-состояния – не менее двух. Таким образом, у человека в такой ситуации есть некоторая свобода выбора между несколькими альтернативными вариантами. Каждый вариант выбора (выбор альтернативы) приводит к результату, который называется исходом. У человека есть свои представления о достоинствах и недостатках отдельных исходов, свое собственное отношение к ним, а, следовательно, и к вариантам решения. Таким образом, у человека, принимающего решение, есть система предпочтений.

Под принятием решений понимается выбор наиболее предпочтительного решения из множества допустимых альтернатив.

В общем случае процесс принятия решений включает в себя два этапа: подготовительный и деловой. На первом этапе формализуется и решается задача, а на втором результат предъявляется ЛПР, который одобряет его или отвергает. Таким образом, процесс принятия решений может быть циклическим, поэтому важно, чтобы само ЛПР владел методом и мог сам поставить задачу, либо аналитик, который работает с задачей, был «в команде» и понимал суть решаемой проблемы.

Обычно активные субъекты, которые участвуют в процессе – ЛПР и его контрагенты, имеют различные интересы и стремятся воздействовать на процесс принятия решения (ППР) в своих целях. Это может выражаться в сокрытии истинного мнения и намерений при принятии решения, искажении информации и т.п. Такое поведение участников может привести к решению, далекому от оптимального или справедливого.

Участники ППР должны в общем случае обладать: памятью (способностью накапливать информацию), способностью к прогнозу (могут использовать информацию для предвидения результатов решения), индивидуальными предпочтениями (различные результаты оценивают по разному), могут быть благожелательны (из двух равных для себя решений субъект может выбрать тот, который устроит противника).

Основополагающий принцип ТПР сформулировали Нейман и Моргенштерн: лицо, принимающее решение, должно всегда выбирать альтернативу с максимально ожидаемой полезностью. Этот результат строится на ряде аксиом, его называют гипотезой ожидаемой полезности. Поэтому и задачи формулируются соответственным образом: чем полезнее, предпочтительнее альтернатива - тем выше численная оценка – «чем больше, тем лучше».

В самом общем случае первый этап решения задачи ТПР предполагает установление всех возможных способов действия (альтернативы), их последовательности и числовой оценки, целей участников процесса принятия решений, природы влияния на этот процесс различных случайных и детерминированных управляющих факторов.

Затем подбирается соответствующая модель и метод решения задачи. На сегодняшний день теория достигла состояния, когда разработаны модели для описания практически всех задач принятия решений.

В рамках современной ТПР разработаны модели для описания практически всех типов задач принятия решений, каждому из которых отвечают определенные аналитические методы. Существует довольно много подходов к классификации задач теории принятия решений. Наиболее общий из них предполагает выделение классов на основе следующих классификационных признаков:

– по характеру учета времени: статические и динамические;

– по количеству целей исследования: одна или несколько;

– по количеству критериев: один или несколько;

– по структуре участников: с одним участником, двумя, конечным числом и бесконечным;

– по характеру исходных данных: детерминированные и стохастические и т.д.

Каждому классу задач соответствуют методы ТПР: линейное и нелинейное программирование, критериальный анализ, теория игр и вариационных рядов. Все эти классификации верны, но охватывают неравноценные области проблем, многие из дисциплин перекрывают друг друга по постановке задач и методам решения.

6.2. Эволюция теории принятия решений. ЭВМ в принятии решений

В своем развитии теория принятия решений прошла через три стадии.

На первой стадии развивался дескриптивный подход к принятию решений. Здесь усилия ученых были направлены на описание процесса выбора решений человеком в целях определения рационального зерна, характерного для всякого разумного выбора. В результате проведенных исследований оказалось, что большинство людей действуют интуитивно, проявляя при этом непоследовательность и противоречивость в своих суждениях. Положительным аспектом исследований в области дескриптивного подхода явилось то, что удалось дать достаточно четкий ответ на вопрос, что может и чего не может человек, решая задачу выбора.

На второй стадии исследователи разрабатывали нормативный подход к принятию решений. Однако и здесь их постигла неудача, поскольку идеализированные теории, рассчитанные на сверхрационального человека с мощным интеллектом, не нашли практического применения.

На третьей стадии был развит прескриптивный подход к принятию решений. Он оказался наиболее плодотворным, поскольку предписывал, как должен поступать человек с нормальным интеллектом, желающий напряженно и систематизировано обдумывать все аспекты своей задачи. Прескриптивный подход не гарантирует нахождения оптимального решения в любой ситуации, но обеспечивает выбор такого решения, которое не обременено противоречиями и непоследовательностями.

Данный подход предъявляет к человеку серьезные требования по освоению методов и приемов теории принятия решений, а также предписывает проведение многочисленных вычислений, связанных с реализацией этих методов.

Первоначальным импульсом для применения ЭВМ в процессе принятия решений явилась необходимость проведения большого объема вычислений для получения обобщенной оценки путем синтеза всех плюсов и минусов по каждой альтернативе. На этом шаге решением ЗПР занимались специалисты, имеющие широкие знания как в области методов принятия решений, так и в программировании на ЭВМ.

Поскольку на практике указанное сочетание знаний является редким, возникла новая категория специалистов – аналитиков в области принятия решений.

Аналитики владели методами принятия решений и навыками программирования и выступали в роли посредников между ЛПР и ЭВМ. Аналитик выполнял следующие функции: уточнял совместно с ЛПР постановку задачи, выбирал метод принятия решений, адекватный задаче, собирал необходимую статистическую и экспертную информацию, строил модель задачи, организовывал обработку накопленной информации на ЭВМ, представлял полученные результаты ЛПР и их интерпретировал.

Следующий шаг в применении ЭВМ для принятия решений был связан с созданием диалоговых систем, позволявших менять интересующие исследователя параметры заложенной в память ЭВМ модели задачи принятия решений, выбирать алгоритм поиска решения или его параметров, исследовать чувствительность полученного решения. Такие системы позволяли получать исчерпывающую информацию для всестороннего обоснования выбираемых решений.

В настоящее время в связи с возросшими возможностями современных ЭВМ разработаны программные информационные системы, обеспечивающие поддержку процесса принятия решений на всех его фазах. Большинство систем принятия решений реализовано на персональных ЭВМ.

6.3. Схема процесса принятия решений

Общая схема процесса принятия решений включает следующие основные этапы:

Этап 1. Предварительный анализ проблемы.

На этом этапе определяются:

– главные цели;

– уровни рассмотрения, элементы и структура системы (процесса), типы связей;

– подсистемы, используемые ими основные ресурсы и критерии качества функционирования подсистем;

– основные противоречия, узкие места и ограничения.

Этап 2. Постановка задачи.

Постановка конкретной ЗПР включает:

– формулирование задачи;

– определение типа задачи;

– определение множества альтернативных вариантов и основных критериев для выбора из них наилучших;

– выбор метода решения ЗПР.

Этап 3. Получение исходных данных.

На данном этапе устанавливаются способы измерения альтернатив. Это либо сбор количественных (статистических) данных, либо методы математического или имитационного моделирования, либо методы экспертной оценки. В последнем случае необходимо решить задачи формирования группы экспертов, проведения экспертных опросов, предварительного анализа экспертных оценок.

Этап 4. Решение ЗПР с привлечением математических методов и вычислительной техники, экспертов и лица, принимающего решение.

На этом этапе производятся математическая обработка исходной информации, ее уточнение и модификация в случае необходимости. Обработка информации может оказаться достаточно трудоемкой, при этом может возникнуть необходимость совершения нескольких итераций и желание применить различные методы для решения задачи. Поэтому именно на этом этапе возникает потребность в компьютерной поддержке процесса принятия решений, которая выполняется с помощью автоматизированных систем принятия решений.

Этап 5. Анализ и интерпретация полученных результатов.

Полученные результаты могут оказаться неудовлетворительными и потребовать изменений в постановке ЗПР. В этом случае необходимо будет возвратиться на этап 2 или этап 1 и пройти заново весь путь. Решение ЗПР может занимать достаточно длительный промежуток времени, в течение которого окружение задачи может измениться и потребовать корректировок в постановке задачи, а также в исходных данных (например, могут появиться новые альтернативы, требующие введения новых критериев). Задачи принятия решений можно разделить на статические и динамические. К первым относятся задачи, которые не требуют многократного решения через короткие интервалы времени. К динамическим относятся ЗПР, которые возникают достаточно часто. Следовательно, итерационный характер процесса принятия решений можно считать закономерным, что подтверждает необходимость создания и использования эффективных систем компьютерной поддержки. ЗПР, требующие одного цикла, можно скорее считать исключением, чем правилом.

6.4. Классификация задач принятия решений

Задачи принятия решений отличаются большим многообразием, классифицировать их можно по различным признакам, характеризующим количество и качество доступной информации. В общем случае задачи принятия решений можно представить следующим набором информации:

$$\langle T, A, K, X, F, G, D \rangle, \tag{6.1}$$

где T – постановка задачи (например, выбрать лучшую альтернативу или упорядочить весь набор);

A – множество допустимых альтернативных вариантов;

K – множество критериев выбора;

X – множество методов измерения предпочтений (например, использование различных шкал);

F – отображение множества допустимых альтернатив в множество критериальных оценок (исходы);

G – система предпочтений эксперта;

D – решающее правило, отражающее систему предпочтений.

Любой из элементов этого набора может служить классификационным признаком принятия решений.

Рассмотрим традиционные классификации:

1. По виду отображения F. Отображение множества A и K может иметь детерминированный характер, вероятностный или неопределенный вид, в соответствии с которым задачи принятия решений можно разделить на задачи в условиях риска и задачи в условиях неопределенности.

2. Мощность множества K. Множество критериев выбора может содержать один элемент или несколько. В соответствии с этим задачи принятия решений можно разделить на задачи со скалярным критерием и задачи с векторным критерием (многокритериальное принятие решений).

3. Тип системы G. Предпочтения могут формироваться одним лицом или коллективом, в зависимости от этого задачи принятия решений можно классифицировать на задачи индивидуального принятия решений и задачи коллективного принятия решений.

Задачи принятия решений в условиях определенности. К этому классу относятся задачи, для решения которых имеется достаточная и достоверная количественная информация. В этом случае с успехом применяются методы математического программирования, суть которых состоит в нахождении оптимальных решений на базе математической модели реального объекта.

Основные условия применимости методов математического программирования следующие:

1. Задача должна быть хорошо формализована, т. е. имеется адекватная математическая модель реального объекта.
2. Существует некоторая единственная целевая функция (критерий оптимизации), позволяющая судить о качестве рассматриваемых альтернативных вариантов.
3. Имеется возможность количественной оценки значений целевой функции.
4. Задача имеет определенные степени свободы (ресурсы оптимизации), т.е. некоторые параметры функционирования системы, которые можно произвольно изменять в некоторых пределах в целях улучшения значений целевой функции.

Задачи в условиях риска. В тех случаях, когда возможные исходы можно описать с помощью некоторого вероятностного распределения, получаем задачи принятия решений в условиях риска. Для построения распределения вероятностей необходимо либо иметь в распоряжении статистические данные, либо привлекать знания экспертов. Обычно для решения задач этого типа применяются методы теории одномерной или многомерной полезности. Эти задачи занимают место на границе между задачами принятия решений в условиях определенности и неопределенности. Для решения этих задач привлекается вся доступная информация (количественная и качественная).

Задачи в условиях неопределенности. Эти задачи имеют место тогда, когда информация, необходимая для принятия решений, является неточной, неполной, неколичественной, а формальные модели исследуемой системы либо слишком сложны, либо отсутствуют. В таких случаях для решения задачи обычно привлекаются знания экспертов. В отличие от подхода, принятого в экспертных системах, для решения ЗПР знания экспертов обычно выражены в виде некоторых количественных данных, называемых предпочтениями.

Выбор и нетривиальность задач принятия решений. Следует отметить, что одним из условий существования задачи принятия решений является наличие нескольких допустимых альтернатив, из которых следует выбрать в некотором смысле лучшую. При наличии одной альтернативы, удовлетворяющей фиксированным условиям или ограничениям, задача принятия решений не имеет места.

Задача принятия решений называется тривиальной, если она характеризуется исключительно одним критерием K и всем альтернативам A_i приписаны конкретные числовые оценки.

Задача принятия решений перестает быть тривиальной даже при одном критерии K, если каждой альтернативе A_i соответствует не точная оценка, а интервал возможных оценок или распределение $f(K/A_i)$ на значениях указанного критерия.

Нетривиальной считается задача при наличии нескольких критериев принятия решений независимо от вида отображения множества альтернатив в множество критериальных оценок их последствий. Следовательно, при наличии ситуации выбора, многокритериальности и осуществлении выбора в условиях неопределенности или риска задача принятия решений является нетривиальной. В табл. 6.1 приведена одна из возможных классификаций, признаками которой являются содержание и тип получаемой экспертной информации.

Таблица 6.1

Классификация методов принятия решений

№ п/п	Содержание информации	Тип информации	Метод принятия решений
1	Экспертная информация не требуется		Метод доминирования Метод на основе глобальных критериев
2	Информация о предпочтениях на множестве критериев	Качественная информация Количественная оценка предпочтительности критериев Количественная информация о замещениях	Лексикографическое упорядочение Сравнение разностей критериальных оценок Метод припасовывания Методы «эффективность-стоимость» Методы свертки на иерархии критериев Методы порогов Методы идеальной точки Метод кривых безразличия Методы теории ценности

Окончание табл. 6.1

№ п/п	Содержание информации	Тип информации	Метод принятия решений
3	Информация о предпочтительности альтернатив	Оценка предпочтительности парных сравнений	Методы математического программирования Линейная и нелинейная свертка при интерактивном способе определения ее параметров
4	Информация о предпочтениях на множестве критериев и о последствиях альтернатив	Отсутствие информации о предпочтениях; количественная и/или интервальная информация о последствиях Качественная информация о предпочтениях и количественная о последствиях Качественная (порядковая) информация о предпочтениях и последствиях Количественная информация о предпочтениях и последствиях	Методы с дискретизацией неопределенности Стохастическое доминирование Методы принятия решений в условиях риска и неопределенности на основе глобальных критериев Метод анализа иерархий Методы теории нечетких множеств Метод практического принятия решений Методы выбора статистически ненадежных решений Методы кривых безразличия для принятия решений в условиях риска и неопределенности Методы деревьев решений Декомпозиционные методы теории ожидаемой полезности

Используемый принцип классификации позволяет достаточно четко выделить четыре большие группы методов, причем три группы относятся к принятию решений в условиях определенности, а четвертая – к принятию решений в условиях неопределенности. Из множества известных методов и подходов к принятию решений наибольший интерес представляют те, которые дают возможность учитывать многокритериальность и неопределенность, а также позволяют осуществлять выбор решений из множеств альтернатив различного типа при наличии критериев, имеющих разные типы шкал измерения (эти методы относятся к четвертой группе).

В свою очередь, среди методов, образующих четвертую группу, наиболее перспективными являются декомпозиционные методы теории ожидаемой полезности, методы анализа иерархий и теории нечетких множеств. Данный выбор определен тем, что эти методы в наибольшей степени удовлетворяют требованиям универсальности, учета многокритериальности выбора в условиях неопределенности из дискретного или непрерывного множества альтернатив, простоты подготовки и переработки экспертной информации.

Охарактеризовать достаточно полно все методы принятия решений, относящиеся к четвертой группе, представляется весьма объемным и самостоятельным исследованием, поэтому в дальнейшем рассматриваются только три подхода к принятию решений в условиях неопределенности, которые получили наиболее широкое воплощение в системах компьютерной поддержки, а именно: подходы, основанные на методах теории полезности, анализа иерархий и теории нечетких множеств.

6.5. Характеристика методов теории полезности

Декомпозиционные методы теории ожидаемой полезности получили наиболее широкое распространение среди группы аксиоматических методов принятия решений в условиях риска и неопределенности.

Основная идея этой теории состоит в получении количественных оценок полезности возможных исходов, которые являются следствиями процессов принятия решений. В дальнейшем на основании этих оценок можно выбрать наилучший исход. Для получения оценок полезности необходимо иметь информацию о предпочтениях лица, ответственного за принимаемое решение.

Парадигма анализа решения может быть сведена к процессу, включающему пять этапов.

Этап 1. Предварительный анализ. На этом этапе формулируется проблема и определяются возможные варианты действий, которые можно предпринять в процессе ее решения.

Этап 2. Структурный анализ. Этот этап предусматривает структуризацию проблемы на качественном уровне, на котором ЛПР намечает основные шаги процесса принятия решений и пытается упорядочить их в виде некоторой последовательности. Для этой цели строится дерево решений (рис. 6.1).

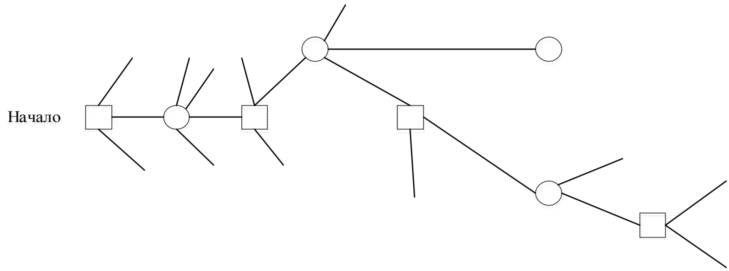


Рис. 6.1. Фрагмент дерева решений

Дерево решений имеет два типа вершин: вершины-решения (обозначены квадратиками) и вершины-случаи (обозначены кружочками). В вершинах-решениях выбор полностью зависит от ЛПР, в вершинах-случаях ЛПР не полностью контролирует выбор, так как случайные события можно предвидеть лишь с некоторой вероятностью.

Этап 3. Анализ неопределенности. На этом этапе ЛПР устанавливает значения вероятности для тех ветвей на дереве решений, которые начинаются в вершинах-случаях. При этом полученные значения вероятностей подлежат проверке на наличие внутренней согласованности.

Для получения значений вероятности привлекается вся доступная информация: статистические данные, результаты моделирования, экспертная информация и т. д.

Этап 4. Анализ полезности. На данном этапе следует получить количественные оценки полезности последствий (исходов), связанных с реализацией того или иного пути на дереве решений.

Исходы (последствия принимаемых решений) оцениваются с помощью функции полезности фон Неймана – Моргенштерна, которая каждому исходу ставит в соответствие его полезность. Построение функции полезности осуществляется на основе знаний ЛПР и экспертов.

Этап 5. Процедуры оптимизации. Оптимальная стратегия действий (альтернатива, путь на дереве решений) может быть найдена с помощью вычислений, а именно: максимизации ожидаемой полезности на всем пространстве возможных исходов. Одно из условий постановки задачи оптимизации - наличие адекватной математической модели, которая связывает параметры оптимизации (в данном случае это альтернативные варианты действий) с переменными, входящими в целевую функцию (функция полезности). В методах теории полезности такие модели имеют вероятностный характер и основаны на том, что оценка вероятности ожидаемого исхода может быть использована для введения числовых оценок возможных вероятных распределений на конечном множестве исходов.

Задача выбора наилучшего решения в соответствии с аксиоматикой теории полезности может быть представлена следующим образом:

$$\max_{A_i \in A}\left[\bar{u}(A)=\int_K u(K)\,f(K/A)\,dK\right],$$

где $u(K)$ - многомерная функция полезности;

K – точка в критериальном пространстве;

$f(K/A)$ - функция плотности условного от альтернативы А распределения критериальных оценок.

Построение функций полезности является основной и наиболее трудоемкой процедурой методов теории полезности, после этого с помощью такой функции можно оценить любое количество альтернатив.

Процедура построения функции полезности включает пять шагов.

Шаг 1. Подготовительный. Главная задача здесь – подбор экспертов и разъяснение им того, как следует выражать свои предпочтения.

Шаг 2. Определение вида функции. Функция полезности должна отражать представления ЛПР и экспертов об ожидаемой полезности возможных исходов. Поэтому множество исходов упорядочивается по их предпочтительности, после чего в соответствие каждому возможному исходу необходимо поставить предполагаемое значение ожидаемой полезности. На этом шаге выясняют, является ли функция полезности монотонной, убывающей или возрастающей, отражает ли она склонность, несклонность или безразличие к риску и т. п.

Шаг 3. Установление количественных ограничений. Здесь определяется интервал изменения аргумента функции полезности и устанавливаются значения функции полезности для нескольких контрольных точек.

Шаг 4. Подбор функции полезности. Необходимо выяснить, являются ли согласованными количественные и качественные характеристики, выявленные к данному моменту. Положительный ответ на этот вопрос равнозначен существованию некоторой функции, которая обладает всеми требуемыми свойствами. Если последует отрицательный ответ, то возникает проблема согласования свойств, что предполагает возврат на более ранние шаги.

Шаг 5. Проверка адекватности. Необходимо убедиться в том, что построенная функция полезности действительно полностью соответствует истинным предпочтениям ЛПР. Для этого применяются традиционные методы сравнения расчетных значений с экспериментальными.

Рассмотренная процедура соответствует задаче со скалярной функцией полезности. В общем случае последняя может быть векторной величиной. Это имеет место, когда ожидаемую полезность невозможно представить единственной количественной характеристикой (задача со многими критериями). Обычно многомерная функция полезности представляется как аддитивная или мультипликативная функция частных полезностей. Процедура построения многомерной функции полезности еще более трудоемка, чем одномерной.

Таким образом, методы теории полезности занимают промежуточное место между методами принятия решений в условиях определенности и методами, направленными на выбор альтернатив в условиях неопределенности. Для применения этих методов необходимо иметь количественную зависимость между исходами и альтернативами, а также экспертную информацию для построения функции полезности. Эти условия выполняются не всегда, что накладывает ограничение на применение методов теории полезности. К тому же следует помнить, что процедура построения функции полезности трудоемка и плохо формализуема.

В настоящее время методы теории полезности достаточно хорошо освещены в научной и учебной литературе. Особого внимания заслуживают работы ученых: А.М. Дуброва, Б.А. Лагоши, Е.Ю. Хрусталева, а также Н.В. Князевского и В.С.Князевской. На основе описанных ими методов реализованы разнообразные компьютерные системы. Наибольшую популярность приобрела промышленная диалоговая система «Альтернатива-Ф», реализующая методы теории полезности и обеспечивающая решение задач многокритериального выбора в условиях определенности, риска и неопределенности.

С учетом сказанного в книге описаны наиболее универсальные и менее освещенные подходы к принятию решений в условиях неопределенности. Наиболее подробно будут рассмотрены автоматизированные методы анализа иерархий и теории нечетких множеств.

6.6. Классификация управленческих решений

6.6.1. Классификация решений по субъектно-объектному признаку

Ведущее место среди субъектов управленческих решений занимает государство.

Решения, принимаемые государством, охватывают все общество в целом, все его сферы и регулируют поведение всех без исключения классов, социальных слоев, групп и отдельных граждан.

Законы выступают в качестве высшей формы права, обладают наивысшей силой и регулируют наиболее важные общественные отношения.

Акты верховных и местных выборных органов – основа для правотворческой деятельности всех органов и организаций.

Акты органов государственного управления – это акты, принимаемые ими в процессе исполнительно-распорядительской деятельности на основе и во исполнение законов и иных актов представительных органов, организаций, должностных лиц и граждан, а также на возникновение, изменение и прекращение конкретных административно-правовых и иных отношений.

Решения, принимаемые субъектом, который одновременно выступает и как объект. Сюда относятся решения, принимаемые на началах общественного самоуправления. В настоящее время этот процесс интенсивно развивается, особенно в связи с акционированием и общим повышением роли производственного коллектива.

6.6.2. Классификация решений по степени определенности ситуации

Определенность. Решение принимается в условиях определенности, когда руководитель в точности знает результат каждого их альтернативных вариантов выбора.

Примером определенного решения является то, что руководитель, по меньшей мере на ближайшую перспективу, точно может установить, какими будут затраты на производство определенного изделия, поскольку арендная плата, стоимость материалов и рабочей силы известны или могу быть рассчитаны с высокой точностью.

Неопределенность. Решения принимаются в условиях неопределенности, когда невозможно оценить вероятность потенциальных результатов. Неопределенность характерна для некоторых решений, которые приходится принимать в быстро меняющихся обстоятельствах.

Сталкиваясь с неопределенностью, руководитель может использовать две основные возможности. Во-первых, попытаться получить дополнительную информацию и еще раз проанализировать проблему. Руководитель сочетает эту дополнительную информацию и анализ с накопленным опытом, способностью к суждению или интуицией, чтобы придать ряду результатов субъективную или предполагаемую вероятность. Вторая возможность – действовать или интуицией и сделать предположение о вероятности событий. Это необходимо, когда не хватает времени на сбор дополнительной информации или затраты на нее чересчур высоки.

Риск. К решениям, принимаемым в условиях риска, относятся такие, результаты которых не являются определенными, но вероятность каждого результата известна. Вероятность определяется как степень возможности свершения данного события и изменяется от 0 до 1. Сумма вероятностей всех альтернатив должна быть равна единице. В условиях определенности существует лишь одна альтернатива.

Наиболее желательный способ определения вероятности – объективность. Вероятность объективна, когда ее можно определить математическими методами или путем статистического анализа накопленного опыта. Например: монета ложится «решкой» в 50% случаев. Вероятность будет определена объективно, если поступит достаточно информации для того, чтобы прогноз оказался статистически достоверным. Например, монета может лечь «орлом» вверх 10,20 и большее число раз в серии.

Во многих случаях организация не располагает достаточной информацией для объективной оценки вероятности, однако, опыт подсказывает, что именно может, скорее всего, случиться с высокой достоверностью. В такой ситуации руководитель может использовать суждение о возможности свершения альтернатив с той или иной субъективной или предполагаемой вероятностью. Например, ставки на скачках, которые делаются до начала забегов.

Люди располагают информацией и опытом – они знают, как выступала лошадь в других соревнованиях – но этого недостаточно для установления объективной вероятности.

6.6.3. Классификация решений по форме

Преобладающей формой являются письменные решения. Это форма решений позволяет внести тот элемент стабильности, упорядоченности и фиксирования информации, без которого немыслимо управление.

Тем не менее, важное место занимают и устные решения, которые в деятельности управленческого и производственного аппарата составляют наиболее оперативную ее часть. Подобные решения могут касаться важных вопросов и должны подкрепляться ответственностью за исполнение.

Еще одной формой решений являются решения, применяющиеся в автоматизированных системах. Это кодированные решения, которые наносятся на специальные документы, различные магнитные носители.

6.6.4. Классификация решений по характеру целей и длительности действий

Стратегические решения. Такие решения обычно касаются коренных проблем. Они принимаются в масштабе объекта управления и выше, рассчитаны на длительный отрезок времени, на решение перспективных задач.

Стратегические цели – это цели, предусматривающие решение масштабных проблем и относятся к компании в целом.

Стратегические решения являются наиболее важными решениями. Они особенно значимы для конкурентоспособности и имеют высокую цену последствий. Такие решения связаны с существенными преобразованиями организации (смена технологии, смена целей, обновление персонала).

Тактические решения. Такие решения, как правило, обеспечивают реализацию стратегических задач. По времени они не превышают одного года.

Тактические цели – это задачи, предусматривающие решение частных проблем, намечаемые менеджерами среднего звена и описывающие шаги, прохождения которых требуют стратегические сели организации.

Оперативные решения. Такие решения связаны с осуществлением текущих целей и задач. По времени они рассчитаны на период, не превышающий месяца.

Операционные цели – это задачи, предусматривающие решения текущих вопросов, намечаемые менеджерами нижнего звена и описывающие действия, необходимые для достижения тактических и стратегических целей.

6.6.5. Классификация решений по их месту и функциям в процессе управления

Оценка обстановки (внешних условий) обычно связывается с подготовкой определенного действия, но в то же время является самостоятельной задачей. Оценить обстановку только путем умозаключения на основании суждений, содержащихся в исходной информации, невозможно. Обычно нет полной гарантии правильного распознавания истинного положения дел и обстоятельств.

Оценка обстановки сама по себе содержит все основные признаки подготовки и принятия решения.

Принятие решения о том, какую информацию следует считать истинной, называют информационным решением. Информационное решение предполагает преобразование информации в такую форму, которая в наибольшей степени соответствует конкретной задаче управления.

Например, руководителю предприятия в течение некоторого времени поступает самая разнообразная информация о состоянии работ на различных участках. В результате обработки этой информации и сопоставления ее с более ранней, руководитель вырабатывает свое представление о производственной ситуации, то есть составляет ее мысленную модель. Это и есть информационное решение.

Следующей разновидностью являются организационные решения. Организационное решение – это выбор альтернатив, который должен сделать руководитель, чтобы выполнить обязанности, обусловленные занимаемой им должностью. Его цель – обеспечение движения к поставленным перед организацией задачам.

Организационное решение состоит в определении структуры, распределении функций между подразделениями и должностными лицами, установлении подчиненности и схемы взаимоотношений.

Особенностью организационных решений является их ориентация на сравнительно широкий диапазон ситуаций. Даже организации разового назначения могут при выполнении поставленной задачи столкнуться с разнообразными условиями. Поэтому их необходимыми качествами являются адаптивность (способность приспосабливаться к обстановке) и устойчивость к посторонним влияниям.

К наиболее сложным и ответственным относят решения, которые называют технологическими или управленческо-технологическими. Класс технологических решений в производственных организациях включает в себя, в частности: определение цели, установление готовности к производству работ и определение их главного направления, распределения сил, средств и способа производства работ, постановку задач подразделениям.

Наиболее ответственным в классе технологических решений является определение цели, на основании чего строятся остальные элементы решения и критерий эффективности. Цель при этом является не внешним фактором по отношению к технологическому решению, а частью его содержания.

В ряде случаев первоначальная цель, хотя бы и четко сформулированная, претерпевает существенные изменения в процессе подготовки технологического решения, появляются дополнительные цели и подцели.

Технологическое решение всегда задает определенное действие, тогда как организационное действие не связано с конкретным действиям, его содержанием и способа осуществления.

6.6.6. Классификация решений по алгоритму

Запрограммированные решения. Запрограммированные решения – это есть результат реализации определенной последовательности шагов или действий, подобных тем, что предпринимаются при решении математического уравнения. Как правили, число возможных альтернатив ограничено, и выбор должен быть сделан в пределах направлений, заданных организацией.

Например, инспектор больницы при составлении графика работы медсестер и санитаров может исходить из формулы, требующей определенного соотношения между числом пациентов и обслуживающего персонала. Если правилами больницы предусмотрена одна медсестра на пять пациентов, то решение принимаются автоматически – на этаже с 50 пациентами нужно иметь 10 сестер.

Руководителю очень важно иметь уверенность в том, что процедура принятия решений, в самом деле, правильна и желательна. Очевидно, если запрограммированная процедура становится неверной и нежелательной, решения, принятые с ее помощью, будут неэффективными, а руководство утратит уважение своих работников и тех людей вне организации, на которых принимаемые решения сказываются. Более того, в высшей степени желательно сообщить об обоснованиях методологии принятия запрограммированных решений тем, кто этой методологией пользуется, нежели просто предложить ее для употребления.

Неспособность ответить на вопросы, начинающиеся с «почему» в связи с процедурой принятия решений, зачастую порождает напряжение и обижает людей, которые должны применять эту процедуру.

Незапрограммированные решения. Решения этого типа требуются в ситуациях, которые в определенной мере новы, внутренне не структурированы или сопряжены с неизвестными факторами. Поскольку заранее невозможно составить конкретную последовательность необходимых шагов, руководитель должен разработать процедуру принятия решения.

К числу незапрограммированных можно отнести решения следующего типа: какими должны быть цели организации, как улучшить продукцию, как усовершенствовать структуру управленческого подразделения, как усилить мотивацию подчиненных. В каждой из подобных ситуаций (как чаще всего бывает с незапрограммированными решениями) истинной причиной проблемы может быть любой из факторов. В то же время, руководитель располагает множеством вариантов выбора.

На практике немногие управленческие решения оказываются запрограммированными или незапрограммированными в чистом виде. Скорее всего, они суть крайние отображения некоторого спектра в случае и с повседневными, и с принципиальными решениями.

Почти все решения оказываются где-нибудь между крайними вариантами. Немногие запрограммированные решения настолько структурированы, что личная инициатива лица, принимающего их, целиком исключается. Даже в ситуации самого сложного выбора методология принятия запрограммированных решений может быть полезна.

6.6.7. Классификация решений по основаниям

Интуитивные решения. Чисто интуитивное решение – это выбор, сделанный только на основе ощущения того, что он правилен. Лицо, принимающее решение, не занимается при этом сознательным взвешиванием «за» и «против» по каждой альтернативе и не нуждается даже, а понимании ситуации. Просто человек делает выбор. То, что мы называем озарением или шестым чувством, и есть интуитивные решения.

Решения, основанные на суждениях. Такие решения иногда кажется интуитивным, поскольку логика их не очевидна. Решение, основанное на суждении, - это выбор, обусловленный знаниями или накопленным опытом. Человек использует знание о том, что случалось в сходных ситуациях ранее, чтобы спрогнозировать результат альтернативных вариантов выбора в существующей ситуации. Опираясь на здравый смысл, он выбирает альтернативу, которая принесла успех в прошлом.

Например: мы делаем выбор, что изучать – программу обучения управлению или программу обучения бухгалтерскому учету, мы, скорее всего, примем решение на основе суждения, исходя из опыта вводных курсов по каждому предмету.

Суждение как основа организационного решения полезно, поскольку многие ситуации в организациях имеют тенденцию к частому повторению. В этом случае ранее принятое решение может сработать снова не хуже, чем прежде.

Поскольку решение на основе суждения принимается в голове руководителя, оно обладает таким значительным достоинством, как быстрота и дешевизна его принятия.

Рациональные решения. Главное различие между решениями рациональным и основанным на суждении заключается в том, что первое не зависит от прошлого опыта.

Рациональное решение обосновывается с помощью объективного аналитического процесса.

6.6.8. Классификация решений по содержанию задачи принятия решений и степени охвата объекта управления

По признаку содержания задачи принятия решений различают: экономические, организационные, технические, технологические, политические.

Сфера их деятельности обусловливает специфические требования к принимаемому решению.

В зависимости от степени охвата объекта, в отношении которого принимается решение, выделяют следующие решения:

– общие, охватывающие весь объект и затрагивающие, как правило, жизненно важные стороны его деятельности;

– частные, касающиеся отдельных сторон деятельности;

– локальные, принимаемые в отношении какого-либо отдельного элемента организационно-экономической системы (например отдела, цеха). Эти решения не затрагивают деятельность всей системы, однако для коллектива той подсистемы, в отношении которой они принимаются, они могут носить характер общих или частных решений.

Классификацию решений в общем виде можно представить в виде табл/ 6.2.

Таблица 6.2

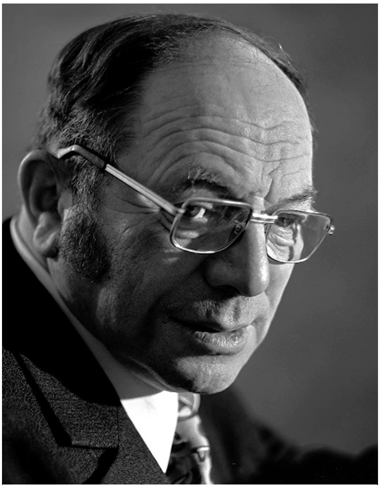
Классификация (типология) управленческих решений

№	Критерий деления	Класс
1	По субъектно-объектному признаку	1.1 Принимаемые государством 1.2 Принимаемые субъектом
2	По степени определенности ситуации	2.1 В условиях определенности 2.2 В условиях неопределенности 2.3 В условиях риска
3	По форме	3.1 Письменные 3.2 Устные 3.3 Кодированные
4	По характеру целей и длительности действий	4.1 Стратегические 4.2 Тактические 4.3 Операционные
5	По месту и функциям в процессе управления	5.1 Информационные 5.2 Организационные 5.3 Технологические (операционно-технологические)
6	По алгоритму	6.1 Запрограммированные 6.2 Незапрограммированные
7	По основаниям	7.1 Интуитивные 7.2 На суждениях 7.3 Рациональные
8	По содержанию задачи принятия решений	8.1 Экономические 8.2 Организационные 8.3 Технические 8.4 Технологические 8.5 Политические
9	По степени охвата объекта управления	9.1 Общие 9.2 Частные 9.3 Локальные

ГЛАВА 7. ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Исследование операций представляет собой научную дисциплину, которая занимается разработкой и применением методов определения оптимальных решений на основе математического, статистического моделирования и различных эвристических подходов в различных областях человеческой деятельности. Ввиду того, что об исследовании операций идет речь в том случае, когда для обоснования принимаемых решений используется тот или иной математический аппарат, содержание данного раздела должно быть весьма обширным.

Традиционно при описании методов исследования операций опираются на математический аппарат следующих разделов: системный анализ, математическое моделирование, линейное, нелинейное, целочисленное и динамическое программирование, теория игр, теория систем массового обслуживания, методы сетевого планирования. В связи с тем, что в данном издании весьма подробно описаны методологические основы системного анализа и математического моделирования, остановимся более подробно на рассмотрении оставшихся разделов математического аппарата исследования операций.



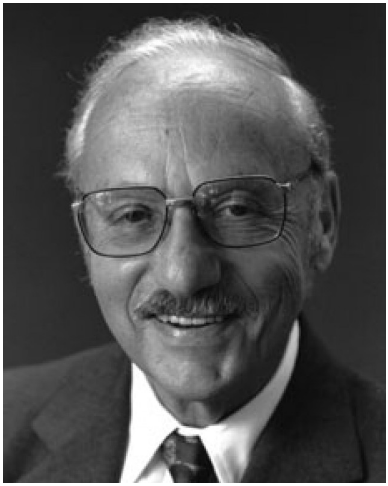
Канторович
Леонид Витальевич
(19.02.1912 – 07.04.1986)

7.1. Основные понятия исследования операций

Одной из первых работ в области исследования операций является «Математические методы организации и планирования производства» Л.В. Канторовича, вышедшая в 1939 г., а в зарубежной литературе – вышедшая в 1947 г. работа Дж. Данцига, посвященная решению экстремальных линейных задач. В 1975 г. Л.В. Канторович стал лауреатом Нобелевской премии за свои работы по оптимальному использованию ресурсов в экономике.

50-ые и последующие годы XX-го столетия были отмечены широким применением в практику полученных фундаментальных теоретических исследований и связанных с этим переосмыслением потенциальных возможностей теории исследования операций.

Важный вклад в развитие новой науки также внесли такие видные ученные, как Дж.Фон. Нейман, Д. Гейл, К. Эрроу, Р. Беллман, Р. Гомори, Е.С. Вентцель, М.К. Гавурин и др. ученые.



Джорж Бернард Данциг
(08.11.1914 – 13.05.2005)

Теория принятия оптимальных решений в наиболее общем смысле представляет собой совокупность математических и численных методов, ориентированных на нахождение наилучших вариантов из множества альтернатив и позволяющих избежать их полного перебора.

Так как размерность практических задач, как правило, достаточно велика, а расчеты в соответствии с алгоритмами оптимизации требуют значительных затрат времени, поэтому методы принятия оптимальных решений ориентированы, главным образом на реализацию их с помощью ЭВМ.

Практическая потребность общества в научных основах принятия решений возникла с развитием науки и техники.

Базовыми понятиями в исследовании операций являются следующие.

Операцией называется всякое мероприятие (система действий), объединенное единым замыслом и направленное к достижению какой-то цели.

Цель исследования операций – предварительное количественное обоснование оптимальных решений.

Решение – всякий определенный выбор зависящих от нас параметров.

Оптимальное решение – это решение, которое по тем или другим признакам предпочтительнее перед другими.

Элементы решения – параметры, совокупность которых образует решение.

Множеством допустимых решений называются заданные условия, которые фиксированы и не могут быть нарушены.

Показатель эффективности – количественная мера, позволяющая сравнивать по эффективности разные решения.

Все решения принимаются всегда на основе информации, которой располагает лицо принимающее решение (ЛПР).

Каждая задача в своей постановке должна отражать структуру и динамику знаний ЛПР о множестве допустимых решений и о показателе эффективности.

Задача называется статической, если принятие решения происходит в наперед известном и не изменяющемся информационном состоянии.

Задача называется динамической – если информационные состояния в ходе принятия решения сменяют друг.

Информационные состояния ЛПР могут по-разному характеризовать его физическое состояние:

– если информационное состояние состоит из единственного физического состояния, то задача называется определенной;

– если информационное состояние содержит несколько физических состояний и ЛПР кроме их множества знает еще и вероятности каждого из этих физических состояний, то задача называется стохастической (частично неопределенной);

– если информационное состояние содержит несколько физических состояний, но ЛПР кроме их множества ничего не знает о вероятности каждого из этих физических состояний, то задача называется *н*еопределенной.

Успешное применение методов принятия решений в значительной мере зависит от профессиональной подготовки специалиста, который должен иметь четкое представление о специфических особенностях изучаемой системы и уметь корректно поставить задачу.

Искусство постановки задач постигается на примерах успешно реализованных разработок и основывается на четком представлении преимуществ, недостатков и специфики различных методов оптимизации.

В первом приближении можно сформулировать следующую последовательность действий, которые составляют содержание процесса постановки задачи:

– установление границы подлежащей оптимизации системы, т.е. представление системы в виде некоторой изолированной части реального мира. Расширение границ системы повышает размерность и сложность многокомпонентной системы и, тем самым, затрудняет ее анализ;

– определение показателя эффективности, на основе которого можно оценить характеристики системы или ее проекта с тем, чтобы выявить "наилучший" проект или множество "наилучших" условий функционирования системы. Обычно выбираются показатели экономического (издержки, прибыль и т.д.) или технологического (производительность, энергоемкость, материалоемкость и т.д.) характера. «Наилучшему» варианту всегда соответствует экстремальное значение показателя эффективности функционирования системы;

– выбор внутрисистемных независимых переменных, которые должны адекватно описывать допустимые проекты или условия функционирования системы и способствовать тому, чтобы все важнейшие экономические решения нашли отражение в формулировке задачи;

– построение модели, которая описывает взаимосвязи между переменными задачи и отражает влияние независимых переменных на значение показателя эффективности. Структура модели, в самом общем случае, включает основные уравнения материальных и энергетических балансов, соотношения, связанные с проектными решениями, уравнения, описывающие физические процессы, протекающие в системе, неравенства, которые определяют область допустимых значений независимых переменных и устанавливают лимиты имеющихся ресурсов. Элементы модели содержат всю информацию, которая обычно используется при расчете проекта. Процесс построения модели является весьма трудоемким и требует четкого понимания специфических особенностей рассматриваемой системы.

Несмотря на то, модели принятия оптимальных решений отличаются универсальностью, их успешное применение зависит от профессиональной подготовки специалиста, который должен иметь полное представление о специфике изучаемой системы.

Основная цель рассмотрения приводимых ниже примеров - продемонстрировать разнообразие постановок оптимизационных задач на основе общности их формы.

Все оптимизационные задачи имеют общую структуру. Их можно классифицировать как задачи минимизации (максимизации) М-векторного векторного показателя эффективности $W_m(x)$, $m = 1, 2, \ldots, M$, N-мерного векторного аргумента $x = (x_1, x_2, \ldots, x_N)$, компоненты которого удовлетворяют системе ограничений-равенств $h_k(x) = 0$, $k = 1, 2, \ldots, K$, ограничений-неравенств $g_j(x) > 0$, $j = 1, 2, \ldots, J$, областным ограничениям $x_{li} < x_i < x_{ui}$, $i = 1, 2, \ldots, N$.

Все задачи принятия оптимальных решений можно классифицировать в соответствии с видом функций и размерностью $W_m(x)$, $h_k(x)$, $g_j(x)$ и размерностью и содержанием вектора x:

– одноцелевое принятие решений – $W_m(x)$ – скаляр;

– многоцелевое принятие решений – $W_m(x)$ – вектор;

– принятие решений в условиях определенности – исходные данные детерминированные;

– принятие решений в условиях неопределенности – исходные данные случайные.

Наиболее разработан и широко используется на практике аппарат одноцелевого принятия решений в условиях определенности, который получил название математического программирования.

В математическом программировании традиционно различают такие классы задач: задачи линейного программирования ($W_m(x)$, $h_k(x)$, $g_j(x)$ – линейны), задачи нелинейного программирования ($W_m(x)$, $h_k(x)$, $g_j(x)$ – нелинейны), задачи целочисленного программирования (значения координат вектора x – целочисленны), задачи динамического программирования ((значения координат вектора x зависят от временного фактора).

Математический аппарат одноцелевого принятия решений в условиях неопределенности представляет собой стохастическое программирование (известны законы распределения случайных величин), теорию игр и статистических решений (закон распределения случайных величин неизвестен).

Эффективность управления зависит от комплексного применения многих факторов и не в последнюю очередь - от процедуры принимаемых решений и их практического воплощения в жизнь. Для того, чтобы управленческое решение было действенным и эффективным, нужно соблюсти определенные методологические основы.

Все методы принятия управленческих решений можно объединить в три группы:

– неформальные (эвристические);

– коллективные;

– количественные.

Неформальные (основанные на аналитических способностях и опыте руководителя) – совокупность логических приемов и методов выбора оптимальных решений руководителем путем теоретического (мыслительного) сравнения альтернатив с учетом накопительного опыта, базирующихся на интуиции. Преимущество заключается в том, что решения, как правило, принимаются оперативно. Недостаток заключается в том, что данный метод базируются, как правило, на интуиции, а отсюда - довольно высокая вероятность ошибок.

Коллективные – метод «мозговой атаки», «мозговой штурм» – применяется, как правило, при необходимости принятия экстренного, сложного, многопланового решения, связанного с экстремальной ситуацией. Это требует от руководителей твердого мышления, умения излагать предложение конструктивно, коммуникабельно, компетентно. В ходе «мозговой атаки» предлагаются различные альтернативы, даже такие, которые выходят за рамки обычных приемов и способов реализации подобных ситуаций в обычных условиях.

Метод Делфи (по названию древнегреческого города Дельфы, известного жившими там мудрецами – предсказателями будущего) – многоуровневое анкетирование. Руководитель объявляет проблему и предоставляет подчиненным возможность формулирования альтернатив. Первый этап формулирования альтернатив проходит без аргументации, т.е. каждым из участников предлагается набор решений. После оценки эксперты предлагают подчиненным рассмотреть данный набор альтернатив. На втором этапе сотрудники должны аргументировать свои предложения, варианты решения. После стабилизации оценок опрос прекращается и принимается предложенное экспертами или скорректированное наиболее оптимальное решение.

Метод «кингисе» – японская кольцевая система принятия решения, суть которой в том, что на рассмотрение готовится проект новации. Он передается для обсуждения лицам по списку, составленному руководителем. Каждый должен рассмотреть предлагаемый проект и дать свои замечания в письменном виде, после чего проводится совещание, на которое приглашаются сотрудники, чье мнение не совсем понятно, либо выходит за рамки обычного решения.

Решения принимаются руководителем на основе экспертных оценок с помощью одного из следующих принципов:

– принципа большинства голосов;

– принципа диктатора – за основу берется мнение одного лица группы;

– принципа Курно – каждый эксперт предлагает свое решение, выбор не должен ущемлять интересов каждого в отдельности;

– принципа Парето – эксперты образуют единое целое, одну коалицию;

– принципа Эджворта – эксперты разбились на несколько групп, каждой из которых невыгодно отменять свое решение. Зная предпочтения коалиций, можно принять оптимальное решение, не нанося ущерба друг другу.

Количественные – в их основе лежит научно-практический подход, предполагающий выбор оптимальных решений путем обработки больших массивов информации.

В зависимости от типа математических функций, лежащих в основе моделей, различают:

– линейное моделирование (используются линейные зависимости);

– динамическое программирование (позволяет вводить дополнительные переменные в процессе решения задач);

– вероятностные и статистические модели (реализуются в методах теории массового обслуживания);

– теорию игр (моделирование таких ситуаций, принятия решения в которых учитывает несовпадение интересов различных подразделений);

– имитационные модели (позволяют экспериментально проверить реализацию решений, изменить исходные предпосылки.

7.2. Линейное программирование

Линейное программирование – наука о методах исследования и отыскания экстремальных (наибольших и наименьших) значений линейной функции, на неизвестные которой наложены линейные ограничения.

Эта линейная функция называется целевой, а ограничения, которые математически записываются в виде уравнений или неравенств, называются системой ограничений.

Математическое выражение целевой функции и ее ограничений называется математической моделью задачи.

В общем виде математическая модель задачи линейного программирования (ЛП) записывается таким образом:

$$Z(x) = C_1X_1 + C_2X_2 + \ldots + C_jX_j + \ldots + C_nX_n \to \max(\min)$$

при ограничениях:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + a_{2n}x_n = b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ij}x_j + \ldots + a_{in}x_n = b_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nj}x_j + \ldots + a_{mn}x_n = b_n, \end{cases}$$

$$x_j \geq 0, \quad i = \overline{1,m}, \quad j = \overline{1,n},$$

где X_i – неизвестные; a_{ij}, b_j, C_i – заданные постоянные величины.

Все или некоторые уравнения системы ограничений могут быть записаны в виде неравенств.

Математическая модель в более краткой записи имеет вид:

$$Z(x) = \sum C_iX_i \to \max(\min),$$

при $$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad x_j \geq 0, \quad i = \overline{1,m}, \quad j = \overline{1,n}.$$

Допустимым решением (планом) задачи линейного программирования называется вектор $X = (x_1, x_2, ,...x_n)$, удовлетворяющий системе ограничений задачи. Множество допустимых решений образует область допустимых решений (ОДР). Допустимое решение, при котором целевая функция достигает своего экстремального значения, называется оптимальным решением задачи линейного программирования и обозначается $X_{опт}$. Допустимое решение (r – ранг системы ограничений):

$$X\left(x_1, x_2, \ldots x_r, 0, \ldots, 0\right)$$

называется базисным и рассматривается как опорное решение при поиске оптимального решения.

Математическая модель задачи ЛП может быть канонической и неканонической. Если все ограничения системы заданы уравнениями и переменные X_j неотрицательны, то такая модель задачи называется канонической. Если хотя бы одно ограничение является неравенством или какая-то переменная может принимать отрицательные значения, то модель задачи ЛП является неканонической. Чтобы перейти от неканонической модели к канонической, необходимо в каждое неравенство ввести балансовую переменную x_{n+1} . Если знак неравенства «$<$», то балансовая переменная вводится со знаком плюс, если знак неравенства «$>$», то – минус. В целевую функцию балансовые переменные не вводятся.

Чтобы составить математическую модель задачи ЛП, необходимо:

– ввести обозначения переменных;

– исходя из цели экономических исследований, составить целевую функцию;

– учитывая ограничения в использовании экономических показателей задачи и их количественные закономерности, записать систему ограничений.

Каждая задача линейного программирования, называемая прямой или исходной, тесно связана с другой задачей, ее называют двойственной.

Математические модели этих задач имеют следующий вид.

прямая задача:

$$Z_{\max} = \sum_{i=1}^{n} c_i \cdot x_i$$

$$\sum_{j=1}^{m} a_{ij} \cdot x_i \le b_j$$

$$где, (j = 1, 2, ..., m)$$

$$x_i \ge 0, где\ (i = 1, 2, ..., n)$$

двойственная задача:

$$Z'_{\min} = \sum_{j=1}^{m} b_j y_j$$

$$\sum_{i=1}^{n} a_{ij} \cdot y_j \ge C_i,$$

$$где \quad (i = 1, 2, ..., n).$$

$$y_j \ge 0 \quad (j = 1, 2, ..., m)$$

Эти задачи экономически могут быть сформулированы следующим образом.

Прямая задача: сколько и какой продукции x_i $\left(i=1,2,\dots,n\right)$ надо произвести, чтобы при заданных стоимостях единицы продукции c_i, объемом имеющихся ресурсов b_j $\left(j=1,2,\dots,m\right)$ и нормах расхода ресурсов a_{ij} максимизировать выпуск продукции в стоимостном виде.

Двойственная задача: какова должна быть оценка единицы каждого ресурса y_j $\left(j=1,2,\dots,m\right)$, чтобы при заданных b_j, c_i и a_{ij} минимизировать общую оценку затрат на все ресурсы.

Правила построения двойственной задачи по прямой задаче:

1. Если прямая задача решается на максимум, то двойственная задача решается на минимум, если прямая задача решается на минимум то двойственная на максимум;
2. В задаче на максимум ограничения-неравенства имеют вид «≤», а в задаче на минимум – «≥»;
3. Каждому ограничению прямой задачи соответствует переменная двойственной задачи, в другой модели ограничению двойственной задачи соответствует переменная прямой задачи;
4. Матрица системы ограничений двойственной задачей получается из матрицы систем ограничений прямой задачи транспонированием;
5. Свободные члены системы ограничений прямой задачи являются коэффициентами при соответствующих переменных целевой функции двойственной задачи и наоборот;
6. Если на переменную прямой задачи наложено условие неотрицательности, то соответствующее ограничение двойственной задачи записывается как ограничение-неравенство, иначе – как ограничение равенство;
7. Если какое либо ограничение прямой задачи записано как равенство, то на соответствующую переменную двойственной задачи условие неотрицательности не налагается.

Пример 7.1.

Прямая задача:	Двойственная задача:
$Z_{\max}=3x_1+5x_2+8x_3$	$Z'_{\min}=12y_1+18y_2+20y_3$
$x_1+2x_2+3x_3\le 12$	$y_1+5y_2+2y_3\ge 3$
$5x_1+x_2+2x_3\le 18$	$2y_1+y_2+3y_3\ge 5$
$2x_1+3x_2+4x_3\le 20$	$3y_1+2y_2+4y_3\ge 8$
$x_1,x_2,x_3\ge 0$	$y_1,y_2,y_3\ge 0$

В этой задаче y_1, y_2, y_3 – предельные оценки стоимости единицы каждого ресурса, целевая функция – оценка стоимости всех ресурсов, а каждое ограничение есть условие, что оценка ресурсов, идущих на производство продукции x_1, x_2, x_3, не менее чем цена единицы продукции.

Взаимосвязь решений прямой и двойственной задач находится из трех теорем двойственности.

Первая теорема двойственности: если одна из двойственных задач имеет оптимальное решение, то и другая задача имеет оптимальное решение, причем экстремальные значения целевых функций совпадают $Z(X) = Z'(Y)$. Если одна из двойственных задач неразрешима вследствие неограниченности целевой функции на множестве допустимых решений, то система ограничений другой задачи противоречива.

Вторая теорема двойственности: для того чтобы план X^* и Y^* пары двойственных задач были оптимальными, необходимо и достаточно выполнение условий:

$$x_i^* \left(\sum_{i=1}^{n} a_{ij} x_j^* - C_i \right) = 0 \; ;$$

$$y_j^* \left(\sum_{j=1}^{m} a_{ij} x_i^* - b_j \right) = 0 \; .$$

Эти условия называются условиями дополняющей нежесткости. Из них следует, что если какое-либо неравенство системы ограничений в одной из задач не обращается в строгое равенство оптимальным планом этой задачи, то соответствующий элемент оптимального плана двойственной задачи должен равняться нулю. Если какой-либо элемент оптимального плана одной из задач положителен, то соответствующее ограничение в двойственной задаче её оптимальным планом должно обращаться в строгое равенство, т.е.

если $\sum_{j=1}^{m} a_{ij} x_i^* < b_j$, то $y_j^* = 0, (j = 1, 2, ..., m)$;

если $y_j^* > 0$, то $\sum_{i=1}^{n} a_{ij} x_i^* = b_j$.

Аналогично,

если $\sum_{i=1}^{n} a_{ij} y_j^* > C_1$, $C_1, (i = 1, 2, ..., n)$, *то* $x_i^* = 0$;

если $x_i^* > 0$, то $\sum_{i=1}^{n} a_{ij} y_j^* = C_i, (i = 1, 2, ..., n)$.

Третья теорема двойственности: двойственные оценки показывают приращение функции цели, вызванное малым изменением свободного члена соответствующего ограничения задачи линейного программирования, т.е.

$$dz(x^*)/db_j = y_j^*, (j = 1, 2, ..., m).$$

В последнем выражении дифференциалы заменим приращениями. Тогда получим выражение:

$$\Delta Z(x^*) = y_j^* \cdot \Delta b_j,$$

$$\text{если } \Delta b_j = 1, \text{ тогда } \Delta Z(x^*) = y_j^*.$$

Содержание третьей теоремы двойственности: двойственная оценка численно равна изменению целевой функции при изменении соответствующего ресурса на единицу. Двойственные оценки y_j часто называются скрытыми теневыми или маргинальными оценками ресурсов.

7.3. Нелинейное программирование

Во многих оптимизационных задачах целевая функция, или функции, задающие ограничения, не являются линейными. Такие задачи называются задачами нелинейного программирования.

Пример простой нелинейной задачи:

Предприятие для производства какого-то продукта расходует два средства в количестве x и y соответственно. Это факторы производства, например, машины и труд, два различных вида сырья и т.п., а x и y – затраты факторов производства.

Факторы производства считаются взаимозаменяемыми. Если это «труд» и «машины», то можно применять такие методы производства, при которых величина затрат в сопоставлении с величиной затрат труда оказывается больше или меньше (производство более или менее трудоемкое).

Объем производства, выраженный в натуральных или стоимостных единицах, является функцией затрат производства $Z = f(x, y)$. Эта зависимость называется производственной функцией.

Совокупные издержки выражаются формулой $c_1x_1 + c_2y_2 = b$.

Требуется при данных совокупных издержках определить количество факторов производства, которое максимизирует объем продукции Z.

Математическая модель задачи:

Определить такие переменные x и y, удовлетворяющие условиям

$$c_1x_1 + c_2y = b, \; x \geq 0, \; y \geq 0,$$

при которых функция $z = f(x, y)$ достигнет максимума.

Ограничения могут отсутствовать. В этом случае производится безусловная оптимизация задачи. Как правило, функция z может иметь произвольный нелинейный вид. В теории нелинейной оптимизации выделяют понятие локального экстремума (локального минимума, локального максимума), глобального экстремума, условного экстремума.

Понятие условного экстремума вводится для случая, когда число переменных n не меньше 2 $(n \geq 2)$.

Разница между глобальным и локальным экстремумами предоставлена на рис. 7.1: Точки A и B являются точками локального экстремума, а точка C является точкой глобального экстремума. В то же время точка B не является точкой глобального минимума.

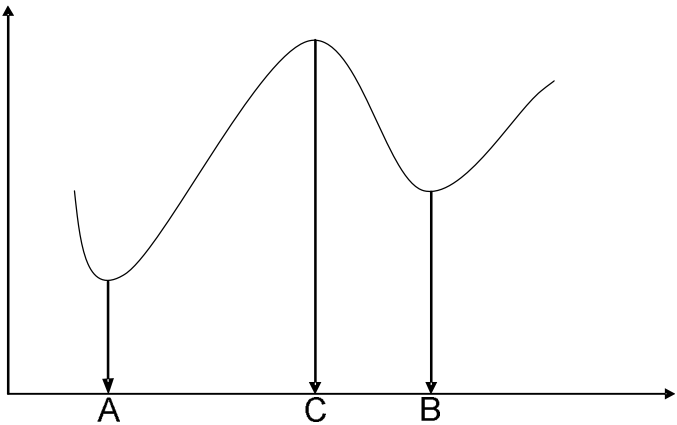


Рис. 7.1. Глобальные и локальные экстремумы

Задачи нелинейного программирования делятся на два класса: имеющие безусловный экстремум и имеющие условный экстремум в зависимости от того есть ли дополнительные условия или нет.

7.4. Целочисленное программирование

Кратко рассмотрим отдельный класс задач целочисленного программирования – задачи целочисленного линейного программирования.

В ряде задач, относящихся к задачам линейного программирования, элементы решения должны выражаться в целых числах. В этих задачах переменные означают количество единиц неделимой продукции.

Задача целочисленного линейного программирования формулируется следующим образом:

Найти такое решение план $X=\left(x_1,x_2,\ldots,x_n\right)$, при котором линейная целевая функция $Z=\sum_{i=1}^{n}c_ix_i$ принимает максимальное или минимальное значение при ограничениях

$$\sum_{j=1}^{m}a_{ij}x_i=b_j, j=1,2,...,m \;;$$

$$x_i>0, i=1,2,...,n\,.$$

$$x_i=\text{целые числа}\,.$$

Задача решается методами линейного программирования.

В случае если переменные оптимального решения оказываются нецелочисленными, то, применяются методы отсечения или метод перебора целочисленных решений.

Метод ветвей и границ заключается в упорядоченном переборе вариантов и рассмотрении лишь тех из них, которые оказываются по определенным признакам перспективными, и отбрасывании бесперспективных вариантов.

Метод ветвей и границ состоит в следующем: множество допустимых решений (планов) некоторым способом разбивается на подмножества, каждое из которых этим же способом снова разбивается на подмножества. Процесс продолжается до тех пор, пока не получено оптимальное целочисленное решение исходной задачи.

Название метода ветвей и границ исходит из того, что в процессе решения задача последовательно «ветвится», заменяясь более простыми. Процесс решения можно продолжать в виде дерева, цифры в узлах (вершинах) которого обозначают план решения задачи (искомые переменные).

При наличии в задаче линейного программирования двух переменных, а в системе ограничения – неравенств, она может быть решена графическим методом без требований целочисленных переменных.

Если оптимальное решение этой задачи является целочисленным, то оно и является оптимальным для исходной задачи.

Если же полученное оптимальное решение не целочисленное, то строится дополнительное линейное ограничение. Оно обладает следующими свойствами:

– оно должно быть линейным;

– должно отсекать найденный оптимальный нецелочисленный план;

– не должно отсекать ни одного целочисленного плана.

Алгоритм графического решения задачи линейного целочисленного программирования:

– построить систему координат $x_1 0 x_2$ и выбрать масштаб;

– найти область допустимых решений (ОДР) системы ограничений задачи;

– построить целевую функцию, являющуюся линией уровня и на ней указать направление нормали;

– переместить линию целевой функции по направлению нормали через ОДР, чтобы она из секущей стала касательной к ОДР и проходила через наиболее удаленную от начала координат точку. Эта точка будет являться точкой экстремума, т.е. решением задачи;

– если окажется, что линия целевой функции параллельна одной из сторон ОДР, то в этом случае экстремум достигается во всех точках соответствующей стороны, а задача линейного программирования будет иметь бесчисленное множество решений;

– найти координаты, точки экстремума и значение целевой функции в ней. Если полученные значения не целочисленные, то перейти к следующему шагу;

– выделить у этих координат область с целочисленными значениями;

– определить новые координаты и построить граф;

– найти точки с целыми значениями искомых переменных, подставить в уравнение целевой функции и найти её значение. Максимальное из полученных значений целевой функции и будет решением задачи.

7.5. Динамическое программирование

Динамическое программирование – раздел оптимального программирования (оптимального управления), в котором процесс принятия решения и управления, может быть разбит на отдельные этапы (шаги).

Динамическое программирование позволяет свести одну сложную задачу со многими переменными ко многим задачам с малым числом переменных. Это значительно сокращает объем вычислений и ускоряет процесс принятия управленческого решения.

Процесс является управляемым, если можно влиять на ход его развития.

Управление – совокупность решений, принимаемых на каждом этапе для влияния на ход развития процесса.

Операция – управляемый процесс, т.е. мы можем выбирать какие-то параметры, влияющие на ход процесса и управлять шагами операции, обеспечивать выигрыши на каждом шаге и в целом за операцию.

Решение на каждом шаге называется «шаговым управлением».

Совокупность всех шаговых управлений представляет собой управление операцией в целом.

Математическая постановка задачи динамического программирования: требуется найти такое управление (x) (x), при котором выигрыш обращался бы в максимум:

$$F(x) = \sum_{i=1}^{m} F_i(x_i) \to \max ,$$

где F – выигрыш за операцию;

$F_i(x_i)$ – выигрыш на i-м шаге;

x – управление операцией в целом;

x_i – управление на i-м шаге $(i = 1, 2, \ldots, m)$. В общем случае шаговые управления $(x_1, x_2, \ldots, x_m)$ могут стать числами, векторами, функциями.

Управление (x^*), при котором достигается максимум, называется оптимальным управлением. Оптимальность управления состоит из совокупности оптимальных шаговых управлений $x^* = x_1^*, x_2^*, \ldots, x_m^*$.

$F^* = \max\{F^*(x^*)\}$ – максимальный выигрыш, который достигается при оптимальном управлении x^*.

Ричард Эрнст Беллман
(26.08.1920 – 19.03.1984)

Исходя из условий каждой конкретной задачи длину шага выбирают таким образом, чтобы на каждом шаге получить простую задачу оптимизации и обеспечить требуемую точность вычислений.

Основным методом динамического программирования является метод рекуррентных соотношений, который основывается на использовании принципа оптимальности, разработанного американским математиком Р.Беллманом.

Суть принципа: каковы бы ни были начальное состояние на любом шаге и управление, выбранное на этом шаге, последующие управления должны выбираться оптимальными относительно состояния, к которому придет система в конце каждого шага.

Использование данного принципа гарантирует, что управление, выбранное на любом шаге, не локально лучше, а лучше с точки зрения процесса в целом. Если S_i – состояние системы на i-м шаге, то схему процесса можно изобразить следующим образом:

Условная оптимизация $\longleftarrow$

$$S_0 \underset{шаг}{\overset{x_1}{\longrightarrow}} S_1 \underset{шаг}{\overset{x_2}{\longrightarrow}} S_2 \underset{шаг}{\overset{x_3}{\longrightarrow}} S_3 \ldots \underset{шаг}{\overset{x_m}{\longrightarrow}} S_m$$

Безусловная оптимизация $\longrightarrow$

Основная рекуррентная формула динамического программирования в случае решения задачи максимизации имеет вид:

$$f_m(i) = \max \left\{ \begin{array}{c} f_m\left(стоимость\ шага\right) + \\ + f_{m+1}\left(новое\ состояние\ перед\ шагом\ m+1\right) \end{array} \right\},$$

где максимум в данной формуле берется по всем возможным решениям в ситуации, когда система на шаге m находится в состоянии i, а величина $f_m\left(i\right)$ – максимальная прибыль завершения задачи из состояния i, если предположить, что на шаге m, система находится в состоянии i.

Максимальная прибыль может быть получена максимизацией суммы прибылей самого шага m и максимальной прибыли шага $\left(m+1\right)$ и далее, чтобы дойти до конца задачи.

Планируя многошаговую операцию надо выбирать управление на каждом шаге с учетом всех его будущих последствий на ещё предстоящих шагах.

Управление на i-м шаге выбирается не так, чтобы выигрыш именно на данном шаге был максимальным, а так, чтобы была максимальна сумма выигрышей на всех оставшихся шагах плюс данный шаг.

Среди всех шагов последний шаг планируется без оглядки на будущее, т.е. чтобы он сам, как таковой принес наибольшую выгоду.

Задача динамического программирования начинает решаться с конца, т.е. с последнего шага. Решается задача в 2 этапа:

1 этап (от конца к началу по шагам): Проводится условная оптимизация, в результате чего находится условные оптимальные управления и условные оптимальные выигрыши по всем шагам процесса.

2 этап (от начала к концу по шагам): Выбираются (прочитываются) уже готовые рекомендации от 1-го шага до последнего и находится безусловное оптимальное управление x^*, равный $x_1^*, x_2^*, \ldots, x_m^*$.

7.6. Теория игр

Теория игр – это математическая теория, исследующая конфликтные ситуации, в которых принятие решений зависит от нескольких участников.

Математическая модель конфликтной ситуации называется игрой. Стороны, участвующие в конфликте – игроки, а исход конфликта – выигрыш (проигрыш). Выигрыш или проигрыш может быть задан количественно.

Игра называется антагонистической или игрой с нулевой суммой, если выигрыш одного из игроков равен проигрышу другого, поэтому для полного «задания» игры достаточно указать величину выигрыша первого игрока.

Стратегией игрока называется совокупность принципов, определяющих выбор его действий при каждом личном ходе в зависимости от сложившейся ситуации.

Для того чтобы найти решение игры, следует для каждого игрока выбрать стратегию, которая удовлетворяет условию оптимальности, т.е. один из игроков должен получать максимальный выигрыш, когда второй игрок придерживается своей стратегии. В тоже время второй игрок должен иметь минимальный проигрыш, если первый придерживается своей стратегии.

Такие стратегии называются оптимальными.

При выборе оптимальной стратегии следует полагать, что оба игрока ведут себя разумно с точки зрения своих интересов.

Матрица, элементы которой характеризуют прибыль первого игрока при всех возможных стратегиях (обозначается $\left(\alpha_{ij}\right)$), называется платежной матрицей игры.

Величина $\alpha_{ij} = \max\min a_{ij}$ называется нижней ценой игры.

Величина $\beta_{ij} = \min\max a_{ij}$ называется верхней ценой игры.

В некоторых задачах, приводящихся к игровым, имеется неопределенность, вызванная отсутствием информации об условиях, в которых осуществляется действие (погода, покупательский спрос и т.п.). Эти условия зависят не от сознательных действий другого игрока, а от объективной действительности. Такие игры называются играми с природой.

Человек в играх с природой старается действовать осмотрительно, второй игрок (природа и т.п.) действует случайно.

При решении задач, относящихся к теории игр, необходимо правильно классифицировать задачу, потому что методы, применяемые к антагонистическим играм, кардинально отличаются от методов решения игр с природой.

Прежде всего, надо уметь находить верхнюю и нижнюю цены игры, т.к. достаточно много игр решается в чистых стратегиях.

Например, необходимо определить нижнюю и верхнюю цены игры для матрицы, приведенной ниже:

A_i	B_j			α_i $\alpha = \max \alpha_i$
	B_1	B_2	B_3	
A_1	0.4	0.6	0.8	0.4
A_2	1.1	**0.7**	0.9	**0.7**
A_3	0.7	0.3	0.5	0.3
β_j $\beta = \min \beta_j$	1.1	**0.7**	0.9	

Очевидно, что $\alpha = \beta = 0,7 = \left(A2, B2\right)$.

Общее значение нижней и верхней цены игры $\alpha = \beta = \nu$ называется чистой ценой игры. Седловой точке соответствует пара минимаксных стратегий, эти стратегии называются оптимальными, а их совокупность – решением игры.

Если седловой точки нет, то можно применить графический способ или составить модель и решить задачу симплекс-методом.

Геометрический способ решения игр с нулевой суммой применяется к играм, где хотя бы у одного игрока только две стратегии. Иногда возможно упростить игры, применяя следующие принципы:

1. Игрок А стремится увеличить свой выигрыш, поэтому он не будет использовать стратегии, которые заведомо дают ему меньшие суммы;

2. Игрок В стремится уменьшить свой проигрыш, поэтому он не будет использовать стратегии, которые заведомо отнимают большие суммы.

Например, рассмотрим платежную матрицу

$\cancel{7}$	$\cancel{6}$	$\cancel{5}$	4	2
$\cancel{3}$	$\cancel{4}$	$\cancel{3}$	$\cancel{2}$	$\cancel{3}$
$\cancel{5}$	$\cancel{6}$	$\cancel{6}$	3	5
$\cancel{2}$	$\cancel{3}$	$\cancel{3}$	$\cancel{2}$	$\cancel{4}$

Упростим матрицу, вычеркивая заведомо невыгодные стратегии игроков.

Путем упрощения, ее можно свести к матрице $\left(2 \times 2\right)$

B_j \ A_j	B_1	B_1
A_1	4	2
A_2	3	5

Пусть p_1 – вероятность применения игроком A стратегии A_1;
p_2 – вероятность применения игроком A стратегии A_2.
Так как $p_1 + p_2 = 1$, то $p_2 = 1 - p_1$. Тогда получим:

Чистые стратегии игрока B	Ожидаемые выигрыши игрока A
B_1	$4p_1 + 3p_2 = (4-3)p_1 + 3 = p_1 + 3$
B_2	$2p_1 + 5p_2 = (2-5)p_1 + 5 = -3p_1 + 5$

На оси $0x$ разместим точки $p_1 = 0$ и $p_1 = 1$, через которые проведем прямые, перпендикулярны оси $0x$. Подставляя $p_1 = 0$ и $p_1 = 1$ в выражение $p_1 + 3$, найдем значения, которые отложим на соответствующих перпендикулярных прямых. Соединив эти точки, получим прямую. Аналогично построим прямую для выражения $-3p_1 + 5$ (рис. 7.2).

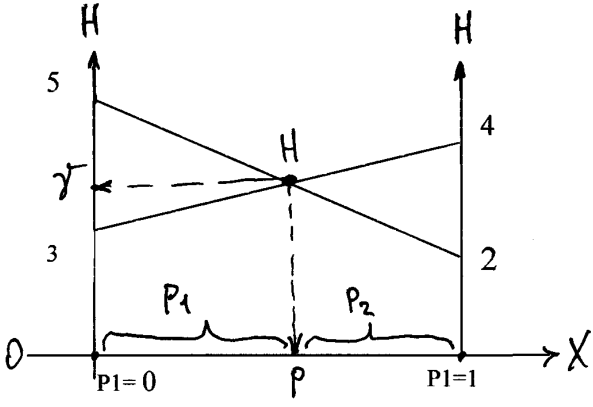


Рис. 7.2. Поиск оптимальных стратегий игроков

Оптимальная стратегия первого игрока найдется из равенства $p_1 + 3$ и $-3p_1 + 5$, т.е. $p_1 = p_2 = 0,5$. $SA = (0,5;0;0,5;0)$, при этом цена игры равна 3,5.

Для второго игрока оптимальная стратегия ищется аналогично.

Если же игра не сводится путем упрощения к $(2\times n)$ или $(m\times 2)$, то составляется математическая модель и задача решается симплекс-методом.

Для того чтобы можно сделать вывод о том какую именно стратегию выбирать игроку при решении игр «с природой», необходимо использовать критерии Вальда, Гурвица, Сэвиджа, Лапласа, Байеса.

1. *Критерий Вальда*. Рекомендуется применять максиминную стратегию. Она достигается из условия $\max\limits_{i}\min\limits_{j}\alpha_{ij}$ и совпадает с нижней ценой игры. Критерий является пессимистическим, считается, что природа будет действовать наихудшим для человека образом, агрессивно, делать все, чтобы помешать нам достигнуть успеха.

2. *Критерий Гурвица* (оптимизма – пессимизма). Критерий рекомендует при выборе решения не руководствоваться ни крайним пессимизмом (всегда рассчитывай на худшее), ни крайним легкомысленным оптимизмом (авось кривая выведет). Критерий рекомендует стратегию, определяемую по формуле

$$H = \max\left\{\gamma \min a_{ij} + (1-\gamma)\max a_{ij}\right\},$$

где γ – степень оптимизма – изменяется в диапазоне $[0,1]$.

Критерий придерживается некоторой промежуточной позиции, учитывающей возможность как наихудшего, так и наилучшего поведения природы. При $\gamma = 1$ критерий превращается в критерий Вальда, при $\gamma = 0$ критерий превращается в критерий максимума. На γ оказывает влияние степень ответственности лица, принимающего решение по выбору стратегии. Чем хуже последствия ошибочных решений, больше желания застраховаться, тем γ ближе к единице.

3. *Критерий Сэвиджа*. Суть критерия состоит в выборе такой стратегии, чтобы не допустить чрезмерно высоких потерь, к которым она может привести. Находится матрица рисков, элементы которой показывают, какой убыток понесет человек, если для каждого состояния природы он не выберет наилучшей стратегии.

Элементы матрицы рисков находится по формуле :

$$r_{ij} = \max a_{ij} - a_{ij},$$

где $\max a_{ij}$ – максимальный элемент в столбце исходной матрицы.

Оптимальная стратегия находится из выражения:

$$H = \min\left\{\max\left(\max a_{ij} - a_{ij}\right)\right\}.$$

4. *Критерий Лапласа.* Этот критерий основывается на принципе недостаточного обоснования. Поскольку вероятности состояния не известны, необходимая информация для вывода, что эти вероятности различны, отсутствует. Поэтому можно предположить, что они равны. Выбор стратегии осуществляется по формуле:

$$H = \max\left\{1/n \sum a_{ij}\right\},$$

где $1/n$ – вероятность реализации одного из состояний.

5. *Критерий Байеса.*

Этот критерий используется для принятия решений в условиях риска. Если в рассмотренных выше критериях, необходимая информация о вероятностях какого-либо состояния отсутствовала, то критерий Байеса действует в условиях не полной информации, т.е. в условиях риска (имеется информация о вероятностях применения стратегий второй стороной). Эти вероятности называются априорными вероятностями.

Выбор стратегии осуществляется по формуле

$$H = \max\left\{\sum p_i a_{ij}\right\}.$$

В условиях полной неопределенности теория не дает однозначных принципов выбора того или иного критерия. Оптимальные стратегии, выбранные по различным критериям, могут быть различны.

Таким образом, окончательный вывод зависит от предпочтений человека, который принимает решение.

7.7. Системы массового обслуживания

Часто приходится сталкиваться с такими ситуациями: очередь покупателей в кассах магазинов; колонна автомобилей, движение которых остановлено светофором; ряд станков, вышедших из строя и ожидающих ремонта, и т.д. Все эти ситуации объединяет то, что системам необходимо пребывать в состоянии ожидания. Ожидание является следствием вероятностного характера возникновения потребностей в обслуживании и разброса показателей обслуживающих систем, которые называют системами массового обслуживания (СМО).

Цель изучения СМО состоит в том, чтобы взять под контроль некоторые характеристики системы, установить зависимость между числом обслуживаемых единиц и качеством обслуживания. Качество обслуживания тем выше, чем больше число обслуживаемых единиц. Но экономически невыгодно иметь лишние обслуживающие единицы.

В промышленности СМО применяются при поступлении сырья, материалов, комплектующих изделий на склад и выдаче их со склада;

обработке широкой номенклатуры деталей на одном и том же оборудовании; организации наладки и ремонта оборудования; определении оптимальной численности обслуживающих отделов и служб предприятий и т.д.

Основными элементами СМО являются: источники заявок; их входящий поток; каналы обслуживания и выходящий поток.

В зависимости от характера формирования очереди СМО различают:

– системы с отказами, в которых при занятости всех каналов обслуживания заявка не встает в очередь и покидает систему необслуженной;

– системы с неограниченными ожиданиями, в которых заявка встает в очередь, если в момент ее поступления все каналы были заняты.

– системы смешанного типа с ожиданием и ограниченной длиной очереди: заявка получает отказ, если приходит в момент, когда все места в очереди заняты. Заявка, попавшая в очередь, обслуживается обязательно.

По числу каналов обслуживания СМО делятся на одноканальные и многоканальные.

В зависимости от расположения источника требований, системы могут быть разомкнутыми (источник заявок находится вне системы) и замкнутыми (источник находится в самой системе).

Рассмотрим в отдельности элементы СМО.

Входящий поток: на практике наиболее распространенным является простейший поток заявок, обладающий свойствами стационарности, ординарности и отсутствия последействия.

Стационарность характеризуется тем, что вероятность поступления определенного количества требований (заявок) в течение некоторого промежутка времени зависит только от длины этого промежутка.

Ординарность потока определяется невозможностью одновременного появления двух или более заявок.

Отсутствие последействия характеризуется тем, что поступление заявки не зависит от того, когда и сколько заявок поступило до этого момента. В этом случае вероятность того, что число заявок, поступивших на обслуживание за промежуток времени t, равно k, определяется по закону Пуассона.

$$P_k(t) = \left((\lambda t)k/k!\right)e^{-\lambda t},$$

где λ – интенсивность потока заявок, т.е. среднее число заявок в единицу времени: $\lambda = 1/\tau$, где τ – среднее значение интервала времени между двумя соседними заявками;

k – число заявок, поступивших на обслуживание за промежуток времени t.

Для такого потока время между двумя соседними заявками распределено экспоненциально с плотностью вероятности: $f(t)=\lambda e^{-\lambda t}$.

Случайное время ожидания в очереди начала обслуживания считают распределенным экспоненциально:

$$f(t)=\upsilon e^{-\upsilon t},$$

где υ – интенсивность движения очереди, т.е. среднее число заявок приходящихся на обслуживание в единицу времени:

$$\upsilon = 1/t_{\text{оч}} ,$$

где $t_{\text{оч}}$ – среднее значение времени ожидания в очереди.

Выходящий поток заявок связан с потоками обслуживания в канале, где длительность обслуживания $t_{\text{обс}}$ является случайной величиной и часто подчиняется показательному закону распределения с плотностью

$$f(t_{\text{обс}})=\mu e^{-\mu t},$$

где μ – интенсивность потока обслуживания, т.е. среднее число заявок, обслуживаемых в ед. времени:

$$\mu = 1/t_{\text{обс}} ,$$

где t – среднее время обслуживания.

Важной характеристикой СМО, объединяющей λ и μ, является интенсивность нагрузки

$$\rho = \lambda/\mu .$$

Рассмотрим n-канальные разомкнутые СМО.

Принцип работы СМО с отказами следующий: заявка, поступившая в систему с отказами и нашедшая все каналы занятыми, получает отказ и покидает систему необслуженной.

Показателем качества обслуживания выступает вероятность получения отказа. Предполагается, что все каналы доступны в равной степени всем заявкам, входящий поток является простейшим, длительность (время) обслуживания одной заявки ($t_{\text{обс}}$) распределена по показательному закону.

Для СМО с отказами вероятность простая каналов обслуживания, когда нет заявок $(k=0)$:

$$P_0 = 1/\left(\sum \rho^k/k!\right).$$

Вероятность отказа в обслуживании, когда поступившая на обслуживание заявка найдет все каналы занятыми $(k = n)$:

$$P_{\text{отк}} = P_n = P_0 \rho^n / n .$$

Вероятность обслуживания:

$$P_{\text{обс}} = 1 - P_{\text{отк}} .$$

Среднее число занятых обслуживанием каналов:

$$n_{\text{з}} = \rho P_{\text{обс}} .$$

Доля каналов, занятых обслуживанием:

$$k_{\text{з}} = n_{\text{з}} / n .$$

Абсолютная пропускная способность СМО:

$$A = \lambda P_{\text{обс}} .$$

Принцип работы СМО с неограниченным ожиданием: заявка, поступившая в систему с неограниченным ожиданием и нашедшая все каналы занятыми, становится в очередь, ожидая освобождения одного из каналов.

Основной характеристикой качества обслуживания является время ожидания (время пребывания заявки в очереди).

Для таких систем характерно отсутствие отказа в обслуживании, следовательно

$$P_{\text{отк}} = 0 \text{ и } P_{\text{обс}} = 1 .$$

Для систем с ожиданием существует дисциплина очереди:

– обслуживание в порядке очереди по принципу «первым пришел – первым обслужен»;

– случайное неорганизованное обслуживание по принципу «последний пришел – первым обслужен»;

– обслуживание с приоритетами по принципу «генералы и полковники вне очереди».

Для СМО с неограниченным ожиданием вероятность простоя каналов, когда нет заявок $(k = 0)$:

$$P_0 = 1 \Big/ \sum \left(\rho^k / k! \right) + \rho^{n+1} \Big/ n! (n - \rho) .$$

При этом предполагается, что $\rho / n < 1$, т.е. интенсивность нагрузки меньше числа каналов.

Вероятность занятости обслуживанием k заявок:

$$P_k = \rho^k \, P_0 / k!,\ 1 \le k \le n\ .$$

Вероятность занятости обслуживанием всех каналов:

$$P_n = P_0 \, \rho^n / n!\ .$$

Вероятность того, что заявка ожидается в очереди:

$$P_{\text{оч}} = \rho^{n+1} / n!(n-\rho)^* \, P_0\ .$$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \rho^{n+1} / (n+\lambda)!(n-\rho)^{2^*} \, P_0\ .$$

Среднее время ожидания заявки в очереди:

$$t_{\text{оч}} = L_{\text{оч}} / \lambda\ .$$

Среднее время ожидания заявки в СМО:

$$t_{\text{смо}} = t_{\text{оч}} + t_{\text{обс}}\ .$$

Среднее число занятых обслуживанием каналов:

$$n_{\text{з}} = \rho\ .$$

Среднее число свободных каналов:

$$n_{\text{св}} = n - n_{\text{з}}\ .$$

Коэффициент занятости каналов обслуживания:

$$k_{\text{з}} = n_{\text{з}} / n\ .$$

Среднее число заявок в СМО:

$$z = L_{\text{оч}} + n_{\text{з}}\ .$$

Принцип работы СМО с ожиданием и с ограниченной длиной очереди: заявка, поступившая в систему с ожиданием с ограниченной длиной очереди и нашедшая все каналы и ограниченную очередь занятыми, покидает систему необслуженной.

Основной характеристикой качества системы является отказ заявке в обслуживании.

Ограничения на длину очереди могут быть из-за:

– ограничения сверх времени пребывания заявки в очереди;

– ограничения сверх длины очереди;

– ограничения общего времени пребывания заявки в системе.

Для СМО с ожиданием и с ограниченной длиной очереди вероятность простоя каналов, когда нет заявок $(k=0)$:

$$P_0 = 1 : \left\{ \sum \rho^k / k! + \rho^{n+1} / n! (n-\rho) \left[1 - (\rho/n)^m \right] \right\},$$

где n – число каналов; m – длина накопителя; ρ – интенсивность нагрузки; k – число заявок, поступивших на обслуживание за промежуток времени t.

Вероятность отказа в обслуживании:

$$P_{\text{отк}} = \rho^{n+m} / n! n^{m*} P_0 .$$

Вероятность обслуживания:

$$P_{\text{обс}} = 1 - P_{\text{отк}} .$$

Абсолютная пропускная способность:

$$A = \lambda P_{\text{обс}} .$$

Среднее число занятых каналов:

$$n_{\text{з}} = A/\mu = \lambda\, P_{\text{обс}} / \mu = \rho P_{\text{обс}} ,$$

где $\rho = \lambda/\mu$

Среднее число заявок в очереди:

$$L_{\text{оч}} = \rho^{n+1} / n^* n! 1 - (\rho/n)^m \left(m + 1 - m\rho/n \right) / \left(1 - \rho/n \right)^{2*} P_0 .$$

Среднее время ожидания обслуживания:

$$t_{\text{оч}} = L_{\text{оч}} / \lambda .$$

Среднее число заявок в системе:

$$z = L_{\text{оч}} + n_{\text{з}} .$$

Среднее время пребывания в системе:

$$t_{\text{смо}} = z/\lambda .$$

7.8. Сетевое планирование

При сетевом планировании определяются оценки продолжительности операций и строится сетевая модель – сетевой график.

Построение сетевого графика позволяет проанализировать все операции и внести улучшения в структуру модели до начала ее реализации.

Календарный сетевой график определяет начало и окончание каждой операции, а также взаимосвязи с другими операциями графика. Он выявляет критические операции, которым надо уделять особое внимание, чтобы закончить все работы в директивный срок.

По выявленным некритическим операциям календарный сетевой график позволяет определить резервы времени, которые можно выгодно использовать при задержке выполнения работ или эффективном использовании трудовых и финансовых ресурсов.

Сетевой график (сетевая модель) – графическое изображение плана выполнения комплекса работ, состоящего из нитей (работ) и узлов (событий), которые отражают логическую взаимосвязь всех операций.

В основе сетевого планирования лежит изображение планируемого комплекса работ в виде графа.

Граф сетевого плана – схема состоящая из заданных точек (вершин), соединенных системой линий.

Сетевой график – это ориентированный граф без контуров (в контуре начальная вершина совпадает с конечной).

Основными элементами сетевых графиков являются: работа, события, путь.

Работа – это активный процесс, требующий затрат ресурсов, либо пассивный (ожидание), приводящий к достижению намеченного результата.

Фиктивная работа – это связь между результатами работ (событиями), не требующая затрат времени и ресурсов, т.е. имеющая нулевую продолжительность.

Событие – это результат выполнения одной или нескольких предшествующих работ.

Путь – любая непрерывная последовательность (цепь) работ и событий.

Критический путь – это путь, который не имеет резервов работы комплекса.

Работы, расположенные на критическом пути, называют *критическими*.

Все остальные работы являются некритическими (ненапряженными) и обладают резервами времени, которые позволяют передвигать сроки их выполнения, не влияя на общую продолжительность выполнения всего комплекса работ.

Ожидание – процесс, требующий затрат времени, но не требующий затрат ресурсов (отдых персонала, ожидание благоприятных условий и т.п.).

Пример сетевого графика показан на рис. 7.3:

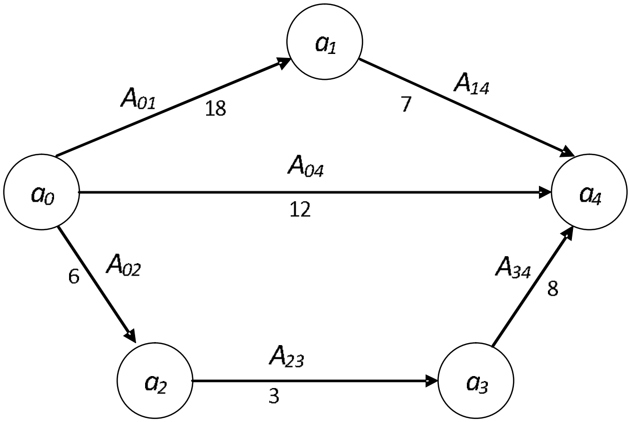


Рис. 7.3. Сетевой график (вариант)

Все работы изображаются на сетевом графике стрелками, величина которых не зависит от продолжительности работы и расхода ресурсов. Стрелки указывают факт и направление движения процесса. Фиктивная работа изображается пунктирной стрелкой. У всех стрелок проставляются индексы, соответствующие наименованию работы, а под ними – время, затрачиваемое на данную работу.

Понятие событие отличается от понятия работы тем, что не является процессом и не связано с затратами времени и ресурсов (разработка сметы закончена, ресурс принят, сборка узла машины завершена). Оно может иметь следующие значения:

– исходное событие, с которого начинаются все работы. В исходное событие не входит ни одна работа (например, получено распоряжение о начале производства продукта).

– завершающее событие – событие, которым заканчивается весь комплекс работ и из него не выходит ни одной работы.

– промежуточные события, или просто события – все события, находящиеся между исходным и завершающим событием.

Любая работа соединяет только 2 события. Событие, из которого выходит работа, является для него начальным или последующим, а куда входит – конечным или последующим.

Работы сетевого графика обозначаются большими буквами и кодируются начальными i и конечными j событиями $\left(A_{04}, A_{01}, A_{23}, \ldots\right)$.

События сетевого графика обозначаются малыми буквами и нумеруются в порядке последовательности развития операции.

Путь в сетевом графике – любая последовательность работ, в которой конечное событие каждой работы является началом следующей за ней работы.

Наибольший по продолжительности путь называется *критическим* и обозначается $L_{\text{кр}}$, а его продолжительность – $T_{\text{кр}}$.

Так, например, на рис. 7.3 критическим является путь

$$a_1 - A_{01} - a_1 - A_{14} - a_4 = 25 \text{ ед. времени.}$$

Выделение критического пути является важнейшим элементом в сетевом планировании.

Критический путь позволяет:

– определить какие работы и события лимитируют выполнение всего комплекса работ;

– позволяет сосредоточить внимание руководителя не на всех работах, а прежде всего – лежащих на критическом пути;

– помогает ускорить выполнение работ за счет привлечения резервов, скрытых в некритических работах.

При описании дуги сетевого графика используют такие величины, необходимые для расчета и получаемые в результате расчета по предшествующим дугам:

i – код начального события работы;

j – код конечного события работы;

ij – код работы (дуги);

t_{ij} – продолжительность работы ij;

$t_i^{\text{р}}$ – ранний срок свершения i-го события, самый ранний срок, в который событие может произойти;

$t_j^{\text{р}}$ – ранний срок свершения j-го события;

$t_i^{\text{п}}$ – поздний срок свершения i-го события,

$t_i^{пред}$ – самый поздний допустимый срок свершения, при котором общая продолжительность работ по графику не увеличится;

$t_j^{\text{п}}$ – поздний срок свершения j-го события;

$t_{ij}^{\text{рн}}$ – раннее начало работы ij;

$t_{ij}^{\text{ро}}$ – раннее окончание работы ij;

$t_{ij}^{\text{пн}}$ – позднее начало работы ij;

$t_{ij}^{\text{по}}$ – позднее окончание работы ij;

$R_{ij}^{\text{п}}$ – полный резерв времени работы, время, на которое можно задержать окончание работы, но так, чтобы при этом общая продолжительность работ по графику не увеличилась;

$R_{ij}^{\text{ч}}$ – частный резерв работы, – время, на которое можно задержать окончание работы так, чтобы ранний срок свершения события j не увеличился.

Рассмотрим алгоритм расчета сетевого графика.

0. Для начального события 1 назначается $t_1^{\text{р}} = 0$.

1. Двигаемся от начального события графика к конечному. Последовательно просматриваются события в порядке возрастания их кодов и вычисляются ранние сроки свершения событий по формуле $t_j^{\text{р}} = \max\left(t_i^{\text{р}} + t_j^{\text{р}}\right)$. Если в событие j входит несколько дуг, то по каждой их них вычисляется величина $t_i^{\text{р}} + t_{ij}$ и в качестве $t_j^{\text{р}}$ принимается большая из рассчитанных величин.

2. Для конечного события графика (код его обозначим k) назначается $t_k^{\text{п}} = t_k^{\text{р}}$ – поздний срок свершения конечного события равен раннему сроку свершения этого события.

3. Двигаемся от конечного события графика к начальному. Просматриваются события в порядке убывания их кодов и вычисляются поздние сроки свершения событий по формуле: $t_i^{\text{п}} = \min\left(t_j^{\text{п}} - t_{ij}\right)$. Если из события i выходит несколько дуг, то по каждой их них вычисляется величина $t_j^{\text{п}} - t_{ij}$ и в качестве $t_j^{\text{п}}$ принимается меньшая. Если расчет произведен без ошибок, то для начального события графика должно оказаться $t_1^{\text{п}} = 0$.

Формулы для вычислений по работам:

$$t_{ij}^{\text{рн}} = t_i^{\text{р}},\ t_{ij}^{\text{по}} = t_j^{\text{п}};$$

$$t_{ij}^{\text{ро}} = t_i^{\text{р}} + t_{ij};\ R_{ij}^{\text{п}} = t_j^{\text{п}} - t_i^{\text{р}} - t_{ij};$$

$$t_{ij}^{\text{пн}} = t_j^{\text{п}} - t_{ij};\ R_{ij}^{\text{ч}} = t_j^{\text{р}} - t_i^{\text{р}} - t_{ij}.$$

Можно ограничиться расчетом на графике.

Иногда результаты расчета показывают в таблице (для каждой работы отводится отдельная строчка):

i	j	t_{ij}	$t_{ij}^{\text{рп}}$	$t_{ij}^{\text{ро}}$	$t_{ij}^{\text{пн}}$	$t_{ij}^{\text{по}}$	$R_{ij}^{\text{п}}$	$R_{ij}^{\text{ч}}$

На рис. 7.4 показан график с рассчитанными сроками свершения событий. Ранние сроки пишутся над событиями, поздние сроки – под событиями. Критический путь показан жирной линией.

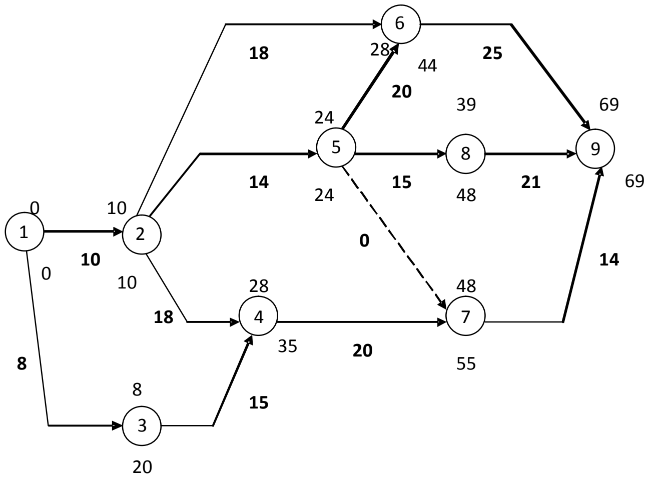


Рис. 7.4. Пример сетевого графика

В табл. 7.1 показаны результаты расчета сетевого графика рис. 7.4:

Таблица 7.1

Результаты расчета сетевого графика

Работа		t_{ij}	Р.Н.	Р.О.	П.Н.	П.О.	Резерв	$R_{ij}^{ч}$
i	j		$t_{ij}^{рп}$	$t_{ij}^{ро}$	$t_{ij}^{пн}$	$t_{ij}^{по}$	$R_{ij}^{п}$	
(1	2)	10	0	10	0	10	0	0
(1	3)	8	0	8	12	20	12	0
(2	4)	18	10	28	17	35	7	0
(2	5)	14	10	24	10	24	0	0
(2	6)	18	10	28	26	44	16	16
(3	4)	15	8	23	20	35	12	5
(4	7)	20	28	48	35	55	7	0
(5	6)	20	24	44	24	44	0	0
(5	7)	0	24	24	55	55	31	24
(5	8)	15	24	39	33	48	9	0
(6	9)	25	44	69	44	69	0	0
(7	9)	14	48	52	55	69	7	7
(8	9)	21	39	60	48	69	9	9

После упорядочения сетевого графика для наглядности рекомендуется дополнить его линейной диаграммой. В ней критическое время комплекса работ равно координате на оси времени самого правого конца всех отрезков диаграммы.

В заключение данного раздела следует отметить, что методы исследования операций, как и любые математические методы, всегда в той или иной мере упрощают, огрубляют задачу, отражая нелинейные процессы линейными моделями, стохастические системы — детерминированными и т.д. Жизнь богаче любой самой сложной схемы. Поэтому не следует ни преувеличивать значения количественных методов исследования операций, ни преуменьшать его, ссылаясь на примеры неудачных решений. Уместно привести в связи с этим известное парадоксальное определение, которое дал крупный американский специалист в этой области Т.А. Саати: «Исследование операций представляет собой искусство давать плохие ответы на те практические вопросы, на которые даются еще худшие ответы другими способами...»

СПИСОК РЕКОМЕНДОВАННОЙ ЛИТЕРАТУРЫ

1. Алгазинов, Е.К. Анализ и компьютерное моделирование информационных процессов и систем / Е.К. Алгазинов, А.А. Сирота. – М.: Диалог МИФИ, 2009. – 414 с.
2. Афонин В.В. Моделирование систем / В.В. Афонин, С.А. Федосин. – М.: БИНОМ, 2010. – 230 с.
3. Будылин А.М. Ряды и интегралы Фурье [Электронный ресурс] / А.М. Будылин. – М., 2002. – Режим доступа: budylin@mph.phys.spbu.ru.
4. Бурков В.Н. Теория графов в управлении организационными системами / В.Н. Бурков, А.А. Заложнев, Д.А. Новиков. – М.: Синтег, 2001. – 124 с.
5. Вентцель Е.С. Теория вероятностей / Е.С. Вентцель. – М.: Наука, 1969. – 576 с.
6. Городецкий А.Я. Информационные системы. Вероятностные модели и статистические решения / А.Я.Городецкий. – СПб: СПбГПУ, 2003. – 326 с.
7. Директор С. Введение в теорию систем / С. Директор, Р. Рорер. – М.: Мир, 1974. – 644 с.
8. Каханер Д., Моулер К., Нэш С. Численные методы и математическое обеспечение / Д. Каханер, К. Моулер, С. Нэш. – М.: Мир, 1998. – 576 с.
9. Клейнрок Л. Теория массового обслуживания / Л. Клейнрок; пер. с англ.. – М.: Машиностроение, 1979. – 432 с.
10. Корнейчук Н.П. Сплайны в теории приближений / Н.П. Корнейчук. – М.: Наука, 1984. – 352 с.
11. Кузнецов О.П. Дискретная математика для инженера / О.П. Кузнецов, Д.М. Адельсон-Вельский. – М: Энергоатомиздат, 1988. – 488 с.
12. Кучук Г.А. Управление ресурсами инфотелекоммуникаций / Г.А. Кучук, Р.П. Гахов, А.А. Пашнев. – М.: Физматлит, 2006. – 220 с.
13. Кучук Г.А. Інформаційні технології управління інтегральними потоками даних в інформаційно-телекомунікаційних мережах систем критичного призначення / Г. А. Кучук. – Х., 2013. – 264 с.
14. Магницкий Н.А. Теория динамического хаоса: монография / Н.А. Магницкий. – М.:УРСС, 2011. – 320 с.
15. Мандельброт Б. Фрактальная геометрия природы / Б. Мандельброт. – М.: Ин-т комп. исследований, 2002. – 656 с.
16. Мендельсон Э. Введение в математическую логику / Э. Мендельсон. – М.: Мир, 1988. – 320 с.

17. Михайлов Г.А. Численное статистическое моделирование / Г.А. Михайлов, .А.В. Войтишек. – М.: Академия, 2006. – 368 с.

18. Наянзин Н.Г. Формирование и оптимизация ресурсных потоков: графокомбинаторные методы и модели. – Владимир, ВГУ, 2006. – 190 с.

19. Оре О. Графы и их применение / О. Оре. – М.: Мир, 1965. – 174 с.

20. Павленко М.А. Методы и процедуры отбора операторов АСУ при использовании интеллектуальных систем поддержки принятия решений / М.А. Павленко // Збірник наукових праць Харківського університету Повітряних Сил. – Х.: ХУПС, 2012. – Вип. 4(33). – С. 171-177.

21. Письменный Д.Т. Конспект лекций по теории вероятностей и математической статистике. – М.: Айрис пресс, 2004. – 208 с.

22. Раскин Л.Г. Многоиндексные задачи линейного программирования (теория, методы, приложения) / Л.Г. Раскин, И.О. Кириченко. – М. : Радио и связь, 1982. – 240 с.

23. Романовский П.И. Ряды Фурье. Теория поля. Аналитические и специальные функции. Преобразование Лапласа / П.И. Романовский. – М.: Наука, 1984. – 304 с.

24. Сигорский В.П. Математический аппарат инженера / В.П. Сигорский. – К.: Техніка, 1977. –766 с.

25. Сирота, А.А. Компьютерное моделирование и оценка эффективности сложных систем / А.А. Сирота. –М.: Техносфера, 2006. –280 с.

26. Скороход А.В. Вероятность. Прикладные аспекты/А.В. Скороход. – М.: МГУ, 2001. – 272 с.

27. Сурмин Ю.П. Теория систем и системный анализ / Ю.П. Сурмин. — К.: МАУП, 2003. — 368 с.

28. Таха Х.А. Введение в исследование операцій / Х.А. Таха. – М.: Вильямс, 2005. — 912 с.

29. Тихонов А.Н. Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. – М.: Наука, 1989. – 344 с.

30. Харари Ф. Теория графов / Ф. Харари. – М.: Мир, 1973. – 428 с.

31. Хинчин А.Я. Работы по математической теории массового обслуживания / А.Я. Хинчин. – М.: Физматгиз, 1963. – 236 с.

32. Цымбал В.П. Теория информация и кодирование / В.П. Цымбал. –К. Вища школа, 1982. – 303 с.

33. Чураков Е.П. Оптимальные и адаптивные системы / Е.П. Чураков. – М.: Энергоатомиздат, 1987. – 242 с.

34. Vapnik V. Estimation of Dependences based on Empirical Data / V. Vapnik. – N.-Y., Berlin : Springer – Verlag, 1987. – 326 p.

Наукове видання
(рос. мовою)

БЕРДНИК Полина Геннадиевна
КУЧУК Георгий Анатольевич
КУЧУК Нина Георгиевна
ОБИДИН Дмитрий Николаевич
ПАВЛЕНКО Максим Анатольевич
ПЕТРОВ Алексей Валерьевич
РУДЕНКО Владислав Николаевич
ТИМОЧКО Александр Иванович

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ЭРГОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ

Монография

Редактор *Сушкова Л.В.*
Техн. редактор *Будулатый В.П.*
Комп'ютерна верстка *Григорук О.Ю.*

Формат 60?84/16 Ум. друк. арк. 14,4 Тираж 00 прим. Зам. № 0301
Свідоцтво держ. реєстру ДК № 977 від 05.07.2002 р.

Видавництво Кіровоградської льотної академії НАУ

м. Кропивницький,
вул. Добровольського, 1,
тел. 39-44-37

Наукове видання
(рос. мовою)

БЕРДНИК Полина Геннадиевна
КУЧУК Георгий Анатольевич
КУЧУК Нина Георгиевна
ОБИДИН Дмитрий Николаевич
ПАВЛЕНКО Максим Анатольевич
ПЕТРОВ Алексей Валерьевич
РУДЕНКО Владислав Николаевич
ТИМОЧКО Александр Иванович

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ЭРГОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ

Монография

Редактор: *Сушкова Л.В.*
Технічний редактор: *Будулатій В.П.*
Комп'ютерна верстка: *Григорук О.Ю.*

Формат 60х84 1/16 Ум. друк. арк. 14,4 Тираж 300 прим. Зам. № 0301
Свідоцтво держ. реєстру ДК № 977 від 05.07.2002 р.

Видавництво Кіровоградської льотної академії НАУ

м. Кропивницький,
вул. Добровольського, 1,
тел.. 39-44-37